PROCEEDINGS

OF THE

AMERICAN PHILOSOPHICAL SOCIETY

HELD AT PHILADELPHIA
FOR PROMOTING USEFUL KNOWLEDGE

VOLUME 81

1939

THE AMERICAN PHILOSOPHICAL SOCIETY
PHILADELPHIA
1939

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS

PUBLICATIONS

OF THE

AMERICAN PHILOSOPHICAL SOCIETY

PROCEEDINGS ($3.00 per volume)

VOLUME 79

VOLUME 80

WILLIAM BELL DINSMOOR. Archæology and Astronomy. pp. 95–173.
January, 1939. $1.00.

No. 2. CHARLES BENEDICT DAVENPORT. Post-Natal Development of the Human Outer Nose. pp. 175–355. 19 pls.
January, 1939. $1.00.

No. 3. EDWIN B. WILSON, CONSTANCE BENNETT, MARGARET ALLEN AND JANE WORCESTER. Measles and Scarlet Fever in Providence, R. I., 1929–1934, with Respect to Age and Size of Family. pp. 357–476.
February, 1939. $0.75.

No. 4. **SYMPOSIUM: INTRINSIC AND ENVIRONMENTAL FACTORS IN AMERICAN POPULATION GROWTH.**

CONRAD TAEUBER. Agriculture and Current Population Trends. pp. 477–489.

WARREN S. THOMPSON. The Social Environment as a Factor in Population Growth. pp. 491–498.

FRANK W. NOTESTEIN. Intrinsic Factors in Population Growth. pp. 499–511.

JOHN C. FLANAGAN. A Study of Psychological Factors Related to Fertility. pp. 513–523.

CLYDE V. KISER. Voluntary and Involuntary Aspects of Childlessness. pp. 525–539.

MORTIMER SPIEGELMAN. Mortality in Relation to Widowhood. pp. 541–558.

WALDEMAR B. KAEMPFFERT. Technological Advance in Relation to Population Trends. pp. 559–568.

CARL CLEVELAND TAYLOR. Prospective Development of Cultural Patterns in Rural America and Their Possible Influence on Population Trends. pp. 569–585.

HARRY L. SHAPIRO. Some Reflections of an Anthropologist on the Future of our Population. pp. 587–600.

ALFRED J. LOTKA. Contact Points of Population Study with Related Branches of Science. pp. 601–626.

February, 1939. $0.75.

VOLUME 81

No. 1. RALPH A. BEEBE. Adsorption Calorimetry on Catalytic Materials and an Account of Some Measurements on Chromic Oxide at – 183° C. pp. 1–14.

NABIH AMIN FARIS. *The Iḥyā' 'Ulūm al-Dīn* of al-Ghazzāli. pp. 15–19.

E. HORNE CRAIGIE. Vascularity in the Brains of Tailed Amphibians. pp. 21–27.

HORACE G. RICHARDS. Land Mollusks from Corn Island, Nicaragua. pp. 29–36.

H. A. SPOEHR AND HAROLD W. MILNER. Starch Dissolution and Amylolytic Activity in Leaves. pp. 37–78.

E. R. FENIMORE JOHNSON. The Preservation and Abuses of Motion Picture Film by Scientific Institutions. pp. 79–92.

M. B. EMENEAU. The Christian Todas. pp. 93–106.

May, 1939. $0.50.

No. 2. **SYMPOSIUM: PROGRESS IN ASTROPHYSICS.**

DONALD H. MENZEL. Problems of the Solar Atmosphere. pp. 107–124.

CHARLOTTE M. SITTERLY. The Composition of the Sun. pp. 125–133.

RUPERT WILDT. The Constitution of the Planets. pp. 135–152.

S. CHANDRASEKHAR. The Internal Constitution of the Stars. pp. 153–187.

CECILIA PAYNE-GAPOSCHKIN. Variable Stars: A Plan of Study. pp. 189–210.

OTTO STRUVE. Stars with Extended Atmospheres. pp. 211–251. 4 pls.

FRED L. WHIPPLE. On the Physical Characteristics and Origin of the Supe- ovæ. pp. 253–264.

DORRIT HOFFLEIT. Observations of Supernovæ. pp. 265–276.

THEODORE DUNHAM, JR. The Material of Interstellar Space. pp. 277–29.

HEN NORRIS RUSSELL. Stellar Energy. pp. 295–307.

Jur ., 1939. $1.00.

No. 3. WILLIAM K. GREGORY AND MILO HELLMAN. On the Evolution and Major Classification of the Civets (Viverridae) and Allied Fossil and Recent Carnivora: A Phylogenetic Study of the Skull and Dentition. pp. 309–392. 6 pls. 1 folder.

CURTIS L. NEWCOMBE AND ANDREW G. LANG. The Distribution of Phosphates in the Chesapeake Bay. pp. 393–420.

D. T. MACDOUGAL. Features of Growth-Control in Trees. pp. 421–445.

ROBLRTS RUGH. Developmental Effects Resulting from Exposure to X-Rays. pp. 447–471. 6 pls.

August, 1939. $1.00.

No. 4. FLETCHER WATSON. A Study of Fireball Radiant Positions. pp. 473–479.

FLETCHER WATSON. Apparent Distribution of Meteoroid Orbits. pp. 481–491.

FLETCHER WATSON. A Study of Telescopic Meteors. pp. 493–504.

J. D. WILLIAMS. Binocular Observations of 718 Meteors. pp. 505–520.

J. D WILLIAMS. Systematic Biases of Meteor Directions. pp. 521–524.

W. HOUSTON. Electrons in Metals. pp. 525–532.

DORRIT HOFFLEIT. On the Necessity for Spectral Sub-classification, with a Discussion of Preliminary Results for 1300 Southern Stars. pp. 533–545.

EI 'ARD W. BERRY. Geology and Paleontology of Lake Tacarigua, :nezuela. pp. 547–568. 7 pls.

E. C. CASE. The Mastodons of Båråboedaer. pp. 569–572. 2 pls.

WILLIAM LYON PHELPS. Notes on Shakespeare. pp. 573–579.

H NRY A. SANDERS. A Soldier's Marriage Certificate in Diploma Form. p. 581–590. 2 pls.

S.ptember, 1939. $0.50.

No. 5. CHARLES WENDELL DAVID. Narratio de Itinere Navali Peregrinorum Hierosolymam Tendentium et Silviam Capientium, A.D. 1189. pp. 591–676.

.ecember, 1939. $0.75.

ADSORPTION CALORIMETRY ON CATALYTIC MATERIALS AND AN ACCOUNT OF SOME MEASUREMENTS ON CHROMIC OXIDE AT —183°C

RALPH A. BEEBE

Moore Laboratory of Chemistry, Amherst College
(Communicated by Hugh S. Taylor)

(*Read Nov. 19, 1938*)

ABSTRACT

I. *Adsorption Calorimetry.* The adsorption of a gas on a solid surface may be: (1) activated adsorption, or (2) van der Waals (physical) adsorption, and only the former of these is in general related to the catalytic activity of the adsorbent material. Because of the great difference in the characteristics of the energy changes occurring during the two types of adsorption, the calorimetric determination of the heats of adsorption provides useful evidence in deciding whether a given process belongs to the activated or the van der Waals type. A vacuum adsorption calorimeter is described which is designed to eliminate the troublesome experimental difficulties arising from the temperature gradients within the instrument.

II. *Measurements at — 183°C.* In the present investigation, heats of adsorption at — 183°C have been measured calorimetrically using the six gases, argon, hydrogen, deuterium, carbon monoxide, nitrogen, and oxygen on a reduced chromic oxide surface. The order of magnitude of the heats indicates that all these gases, with the single exception of argon, are in part held in a state of activated adsorption as well as in the van der Waals state even at the low temperature of — 183°C. Moreover in the experiments with carbon monoxide, nitrogen, and oxygen, the complex form of the time-temperature curves on a partially covered surface has led to the conclusion that the adsorbed gas changes over *on the surface* from an initial van der Waals state to a final state of activated adsorption.

I. ADSORPTION CALORIMETRY

Introduction. As a result of numerous investigations in recent years it is recognized that the activity of so-called contact catalytic agents in accelerating certain chemical reactions is related to the adsorption of one or more of the reactants on the surface of the catalytic material. With the purpose of reaching a better understanding of this relation between the phenomena of adsorption and catalytic activity, a great many investigators have given their attention to the study of adsorption, especially the adsorption of gases on catalytically active metals and metallic oxides. In the course of this development, two types of adsorption have become more or less clearly recognized.* These are: (1) van der Waals (physical) adsorption, and (2) activated adsorption or chemi-

* For references on this subject see Benton and White, *Jour. Am. Chem. Soc.*, 54, 1820 (1932).

sorption.* In general, catalytic activity is associated with activated adsorption only.

Van der Waals adsorption resembles rather closely the condensation of a vapor in equilibrium with its liquid state. As in the case of the liquid-vapor system, equilibrium is always reached rapidly when the gas is brought into contact with the adsorbing surface, and the process of adsorption is easily and completely reversible. The heat liberated during physical adsorption is relatively small, being of the order of magnitude of the heat of vaporization although usually somewhat larger, and, like the latter, it is approximately the same for different gases if they have similar boiling points.

Activated adsorption, on the other hand, may be regarded as a surface reaction in which chemical forces are in operation. As with ordinary chemical reactions, the characteristics of activated adsorption show a wide variation when different specific examples are compared. The process may be fast or slow, it may be reversible or irreversible, it may liberate a large amount of heat or a comparatively small amount depending specifically on the gas and solid surface under consideration. In general, however, the amount of heat liberated is of a higher order of magnitude than the heat of van der Waals adsorption, indicating that a more profound change has occurred during the process of activated adsorption.

It is evident that the thermal data should be very useful in assigning a specific adsorption process to the activated or the van der Waals class. For this reason numerous measurements of heats of adsorption have been made under a wide variety of experimental conditions. Two different general methods have been used: (1) direct calorimetry, and (2) calculation by means of the Clapeyron equation from the experimentally determined equilibrium pressures at two temperatures. The latter method is reliable under favorable circumstances but in many instances there are certain difficulties which make its use unjustifiable or even impossible. We must turn, therefore, to adsorption calorimetry as a check on the latter method and, in some cases, as the only means of obtaining reliable thermal data.

The General Technique of Adsorption Calorimetry. Fig. 1 is a diagram of the apparatus used in the Amherst laboratory for heats of adsorption measurements. The essential parts indicated, are found in any assembly for adsorption calorimetry, although different investiga-

* Many authors now prefer to use the term 'chemisorption' in place of 'activated adsorption.' See Adam, The Physics and Chemistry of Surfaces, Second Edition, p. 265 (1938), (The Oxford University Press). In the present paper the two terms are used synonymously.

tors have used various designs of the gas buret and the manometer and especially of the adsorption calorimeter. After suitable preparation, the catalytic adsorbent material, usually in the form of fine granules, is placed in the calorimeter tube and outgassed at high temperature. The

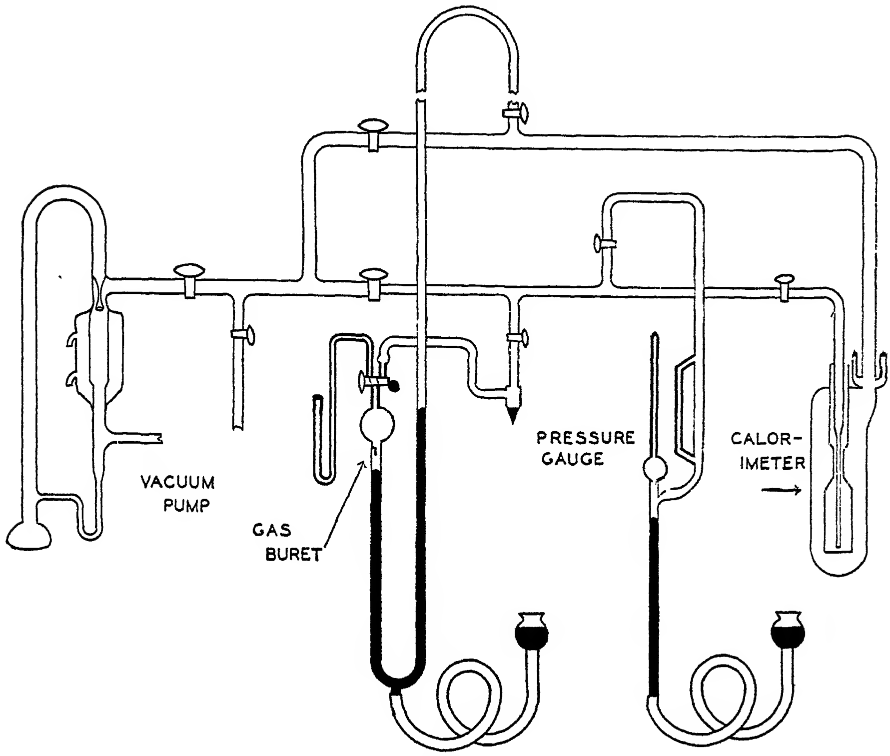


Fig. 1. Adsorption apparatus.

adsorption calorimeter is then brought to the desired temperature, a measured volume of gas is admitted and the amount of heat evolved is determined. From these data the heat of adsorption per mole of gas can be calculated. When sufficient gas is admitted at one time to cover a large fraction of the adsorbing surface, integral heats of adsorption are obtained. More frequently, however, the gas is admitted in successive increments each of which covers only a small fraction of the surface; the heats found by this procedure are called differential heats, although true differential heats would be obtained only when each increment represented an infinitesimal fraction of the total volume adsorbed.

Because of great variation in the characteristics of different specific adsorption reactions, not all of these processes lend themselves equally well to heat measurement. The experimental problem is simplified

if the adsorption is rapid with a resultant rapid evolution of heat, and if a high percentage of the sample of gas is adsorbed, leaving a small correction for unadsorbed gas in the apparatus. Even when these conditions are fulfilled, there are certain difficulties arising in the calorimeter which must be overcome. Unfortunately the significance of these difficulties, which will be discussed later, was not well understood in much of the earlier work, and as a result some of the data in the literature are unreliable.

Types of Calorimeter. The methods of adsorption calorimetry fall into two general classes: (1) those in which the heat evolved produces a physical change, and (2) those in which a temperature rise is produced. In the former class are found: (*a*) the Bunsen ice calorimeter,[1] and (*b*) the liquid oxygen calorimeter of Dewar.[2] In the latter class; (*a*) the adsorption tube may be immersed in a suitable calorimetric liquid bath [3] or (*b*) it may be vacuum jacketed.[4]

The vacuum calorimeter appears to offer the greatest promise because of its sensitivity and because it is the only one which has been used for measurements over a wide range of temperature. In Fig. 2 is shown a simple design of this instrument. The heat capacity of the calorimeter is low, because the latter consists only of the thin walled glass container tube E and the adsorbent material itself. Vacuum calorimeters of this simple type are satisfactory for use under favorable experimental conditions. However, for reliable results under all conditions, it is necessary, because of the experimental difficulties involved, to modify the design of the instrument considerably.

Difficulties Caused by Temperature Gradients. In any calorimetry involving temperature change, the thermometer reading should represent the true average temperature of the whole calorimeter. This specification is not realized in the instrument shown in Fig. 2, owing to temperature gradients set up inside the calorimeter. The causes underlying these temperature gradients have been studied in several different laboratories,[4c, 5] and the results of these investigations have been summarized by Garner and Veal.[4c] The causes appear to be: (1) the loss of heat from the external surface of the calorimeter, (2) the slow rate of distribution of the liberated energy to various parts of the calorimeter, and (3) an unequal distribution of the adsorbed gas throughout the catalyst, the gas being adsorbed on that portion of the adsorbent with which it first comes into contact. These factors are particularly troublesome in the early increments of certain series of differential heat measurements in which no gas remains unadsorbed and the adsorbed gas is unequally distributed. Under these circumstances the heat is liber-

ated locally in one part only, rather than uniformly throughout the adsorbent mass, and this heat will be distributed very slowly across the evacuated spaces between the granules of adsorbent, with the result that the temperature observed, at any specific time after the admission of the gas to the surface, will depend upon the position of the thermocouple relative to the location of that part of the adsorbent mass on which the gas is adsorbed.

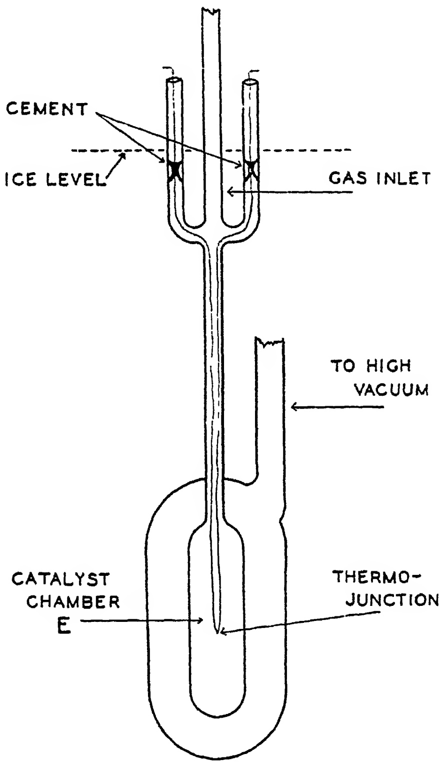


FIG. 2 Simple type vacuum colorimeter.

An Improved Vacuum Calorimeter. Now that the underlying difficulties are understood, it is possible to design a vacuum calorimeter which is practically free from these troublesome temperature gradients. In Fig. 3 is shown the instrument now in use in this laboratory in the measurement of heats of adsorption of gases on chromic oxide and iron. Conditions for rapid heat distribution are improved by using an all-metal calorimeter tube, by mixing copper shot with the adsorbent granules and by inserting six vertical copper vanes not shown in the

diagram.* Especially in the work at − 183°C, the rate of loss of heat from the outside of the calorimeter is very small indeed because of the low rate of radiation, and at 0° the rate of heat loss is still small compared to the rate of distribution throughout the metal calorimeter. Difficulties due to unequal distribution of the adsorbed gas have been minimized by causing the gas to enter through the many fine perfora-

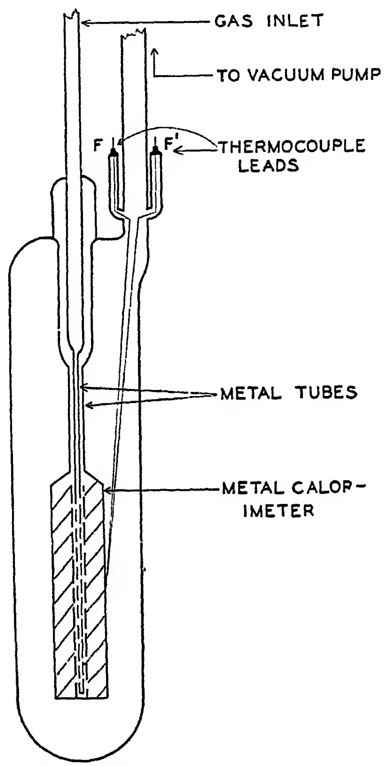


Fig. 3. Metal vacuum calorimeter.

tions in the wall of the central metal tube, thereby bringing it immediately into contact with a large fraction of the adsorbing surface. Test experiments have shown that this improved instrument is reliable even for early increments in which there is no residual gas pressure. The vacuum calorimeter recently developed by Garner and Veal [4c] has been shown by careful tests to give equally trustworthy results.

* It may occur to the reader that a considerable amount of adsorption might happen on the metal surfaces inside the calorimeter. In all probability, at the low pressures employed, the adsorbed molecules are held to the adsorbing surface in a layer only one molecule deep; and a simple calculation will show that the amount of gas adsorbed on the smooth-surfaced materials within the calorimeter, other than the chromic oxide itself, would be negligibly small. Of course the adsorptive capacity of the chromic oxide is increased enormously by its porosity, a large percentage of the adsorption occurring as monolayers on its *internal* surface.

II. Measurements on Chromic Oxide at − 183°C *

In many instances it has been shown that a gas may be held in both the activated and the van der Waals state on the same solid surface. Frequently the two types occur in more or less well separated temperature ranges, the activated type being characteristic of the higher temperature. In certain cases, however, activated adsorption is found even at liquid air temperature occurring concurrently with the low temperature van der Waals process. It was suspected that the adsorption of gases on reduced chromic oxide at − 183°C might be of this complex nature and, in order to identify and unravel the two types of adsorption by means of their characteristic energy changes, the measurement of the heats of adsorption was undertaken.

Apparatus and Materials. The assembled calorimeter (Fig. 3) had a total mass of 103.80 g, and contained 11.95 g of chromic oxide catalyst. A constant temperature of − 183°C was maintained by use of a liquid oxygen bath. The temperature change was measured by a single junction copper-constantan thermocouple and sensitive galvanometer, and the time-temperature curves were recorded photographically. In Fig. 4 is given a typical time-temperature curve for argon at − 183°C. The

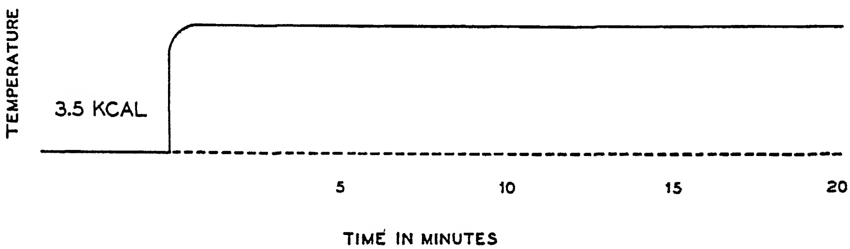


Fig. 4. Time-temperature curve for argon at − 183°.

rapid rise to maximum temperature following the admission of the gas, and the almost negligible rate of cooling after the temperature rise, both indicate that the calorimetric conditions were very satisfactory, and the temperature recorded by the thermocouple was truly the average temperature of the whole metal calorimeter with negligible heat lag due to poor thermal distribution. This type of curve, which was obtained under all conditions in the measurements with argon as well as with the isotopes of hydrogen, will be referred to as a *normal* time-temperature curve in comparison to curves of *abnormal* type resulting from the more complex nature of the adsorption of the other gases studied at − 183°C.

* A more detailed account of these measurements has been published elsewhere [Beebe and Dowden, *Jour. Am. Chem. Soc.*, 60, 2912 (1938)].

The chromic oxide adsorbent was prepared by the method of Burwell and Taylor [6] which involved precipitation of the hydroxide from a chromic nitrate solution and subsequent dehydration of the precipitate followed by reduction in a hydrogen atmosphere at 400°. Following each series of heat measurements, the surface was outgassed for three hours at 400°, except in the case of oxygen. For the removal of the latter gas, the final three hour period of outgassing was preceded by a one hour period of reduction in a hydrogen atmosphere. During this reduction, hydrogen was alternately removed and replaced several times to eliminate the water vapor produced.

The Adsorption Isotherms. It is apparent from Fig. 5 which shows the adsorption isotherms at − 183°C, that a relatively large quantity

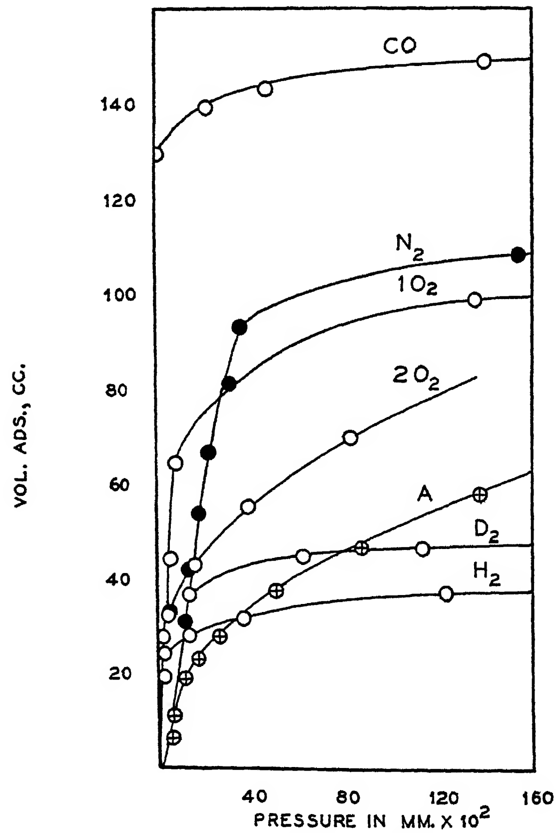


FIG. 5. Adsorption isotherms on 11.95 G Cr_2O_3 at − 183°.

of each of the gases studied is adsorbed with very low residual gas pressure. Indeed, with the carbon monoxide and the hydrogen isotopes,

substantial volumes of gas were taken up at zero pressure. In all these experiments, practically all the adsorption had occurred within one minute after admission to the adsorption tube.

Heat Measurements with Argon and the Hydrogen Isotopes. The results of a series of differential heat measurements at − 183°C with argon are given in the lower curve of Fig. 9. Because of the magnitude of the heats which fall between 4.2 and 2.7 *k*cal./mole, it is concluded that the process for argon is a simple van der Waals adsorption.

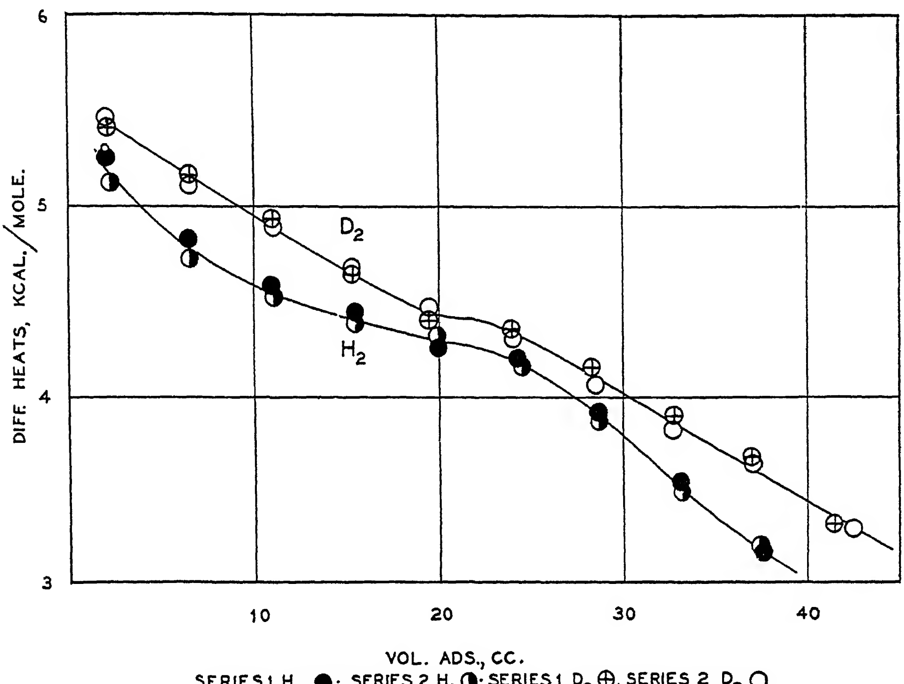


FIG. 6

On the other hand, the results of the heat measurements with hydrogen and deuterium appear to indicate that at least a part of the adsorption of these gases is activated. In no case previously reported has the heat of van der Waals adsorption of hydrogen exceeded 2 *k*cal.; it seems justifiable therefore to assume that a process which produces heats of 3–5 *k*cal. cannot be exclusively of the van der Waals type. This conclusion is consistent with the previously observed [7] catalytic activity of chromic oxide at − 183°C for the reaction $H_2 + D_2 \rightleftharpoons 2HD$, an activity which is probably due to the activated adsorption of both hydrogen isotopes. The differential heats are shown in Fig. 6.

Heat Measurements with Carbon Monoxide. Unlike the gases already discussed, carbon monoxide failed to produce normal time-temperature curves under all conditions. This anomalous behaviour is illustrated by the curves in Fig. 7. Curves I, II and III are the time-temperature curves obtained with approximately equal increments of about 4.5 cc of the gas on a surface to which had been added already 0, 55, and 112 cc respectively, *i.e.* these curves were obtained from measurements on a bare surface, a partly covered surface, and a nearly saturated surface.

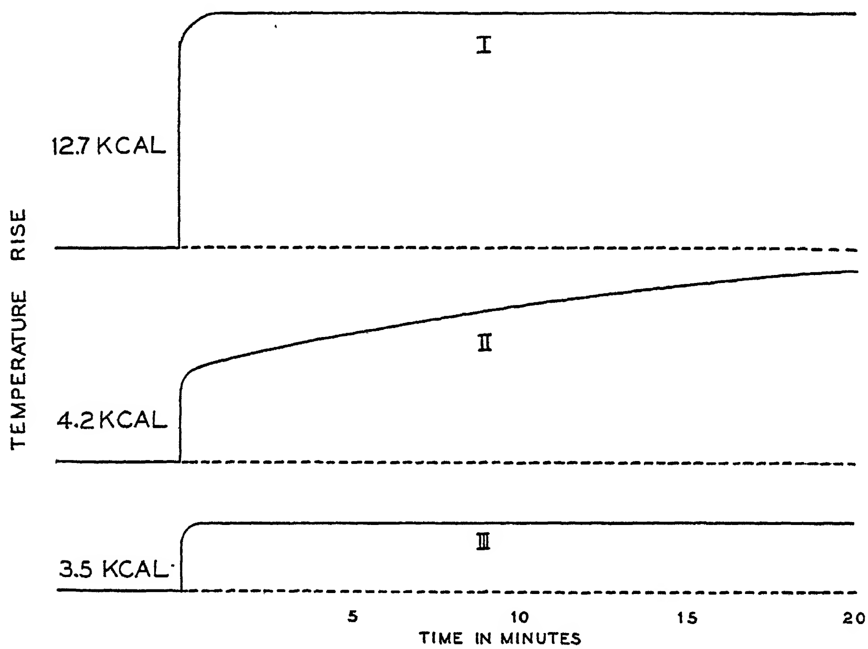
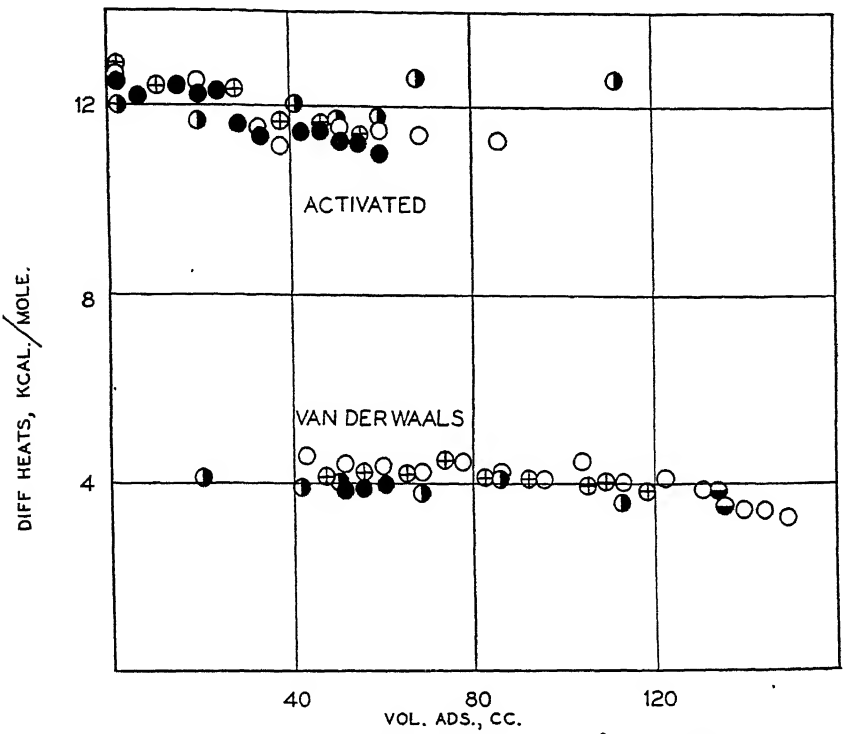


FIG. 7. Time-temperature curves for CO at − 183°.

A normal type of curve was obtained when no carbon monoxide was previously adsorbed (Curve I); however, with successive increments of the gas, the rise to maximum temperature became less and less sharp, and then began to divide into a rapid and a slow stage (Curve II) until finally the slow stage had assumed such a low rate that it could scarcely be detected in the twenty minute period of observation (Curve III). All the other time-temperature curves observed in the series of measurements fitted in form to their intermediate positions in the series. With all the carbon monoxide increments up to 120 cc adsorbed, *all the gas disappeared from the gas phase within one minute after admission.*

From these observations it is concluded that the adsorption of carbon monoxide on chromic oxide at − 183°C consists of two distinct and

separate processes which we shall call Process A and Process B. Process A is rapid for all degrees of surface activity, and gives a heat of approximately 4 *k*cal.; Process B is rapid on a bare surface but becomes progressively slower with successive increments until it is too slow to observe. Process B yields an additional heat of about 8 *k*cal. bringing the total heat up to 12 *k*cal. Process A is doubtless a van der Waals adsorption, and Process B appears to represent a change of the adsorbed



CO ON Cr_2O_3: SERIES 1, 2, AND 3 AT −183° ⊕, ○, AND ●. AFTER OUTGASSING AT −183° ◒; SERIES 1 AT −195° ◑.

Fig. 8

gas on the surface from an initial van der Waals state to a final state of activated adsorption. If this is true the heat values 4 and 12 *k*cal. represent differential heats of van der Waals and of activated adsorption respectively, and have therefore been so designated in Fig. 8. Furthermore, because of the slow evolution of heat (Curve II, Fig. 7) long after all the carbon monoxide had disappeared from the gas phase, it seems probable that the gas underwent this change from the van der Waals to the activated state without leaving the surface.

Experiments with Nitrogen and Oxygen. The heat experiments with nitrogen and oxygen produced the same division into two processes. As in the case of carbon monoxide, the heat of the van der Waals adsorption was about 4 *k*cal. for both these gases. The heats of activated adsorption were however about 8 *k*cal. for nitrogen and 20–27 *k*cal. for oxygen. The differential heats for both activated and van der Waals adsorption are shown in Figs. 9 and 10. The failure of the differential heat data in the activated adsorption of oxygen to coincide in the three experiments performed is due to the impossibility of reproducing the activity of the

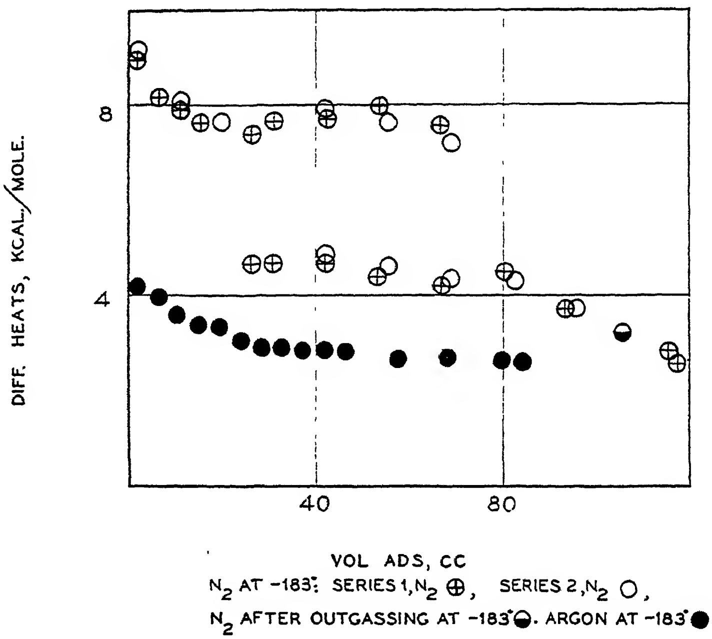


FIG. 9

surface after an oxygen experiment. Unlike the other gases, oxygen is irreversibly adsorbed and can be removed only by reduction in a hydrogen atmosphere, a treatment which alters the activity of the chromic oxide.

Conclusion. It is interesting to compare the data obtained with the four gases, argon, carbon monoxide, nitrogen, and oxygen. Because these substances all boil within the ten degree range − 183°C to − 195°C, we should expect them to have, as they do, approximately the same heats of van der Waals adsorption. The value of 4 *k*cal. which persists for all four gases indicates the non-specific character of the physical adsorption process. On the other hand the heats of the acti-

vated process are 12, 8, and 27 *k*cal. respectively for carbon monoxide, nitrogen and the first oxygen experiment, and they also vary from 27 to 21 *k*cal. in the oxygen experiments on surfaces of different activities. This specificity of the heat evolution to the particular gas and the state of the solid surface is characteristic of activated processes in general.

Attention is called to the novel application of the calorimetric method in the present investigation. The course of the slow process of activated

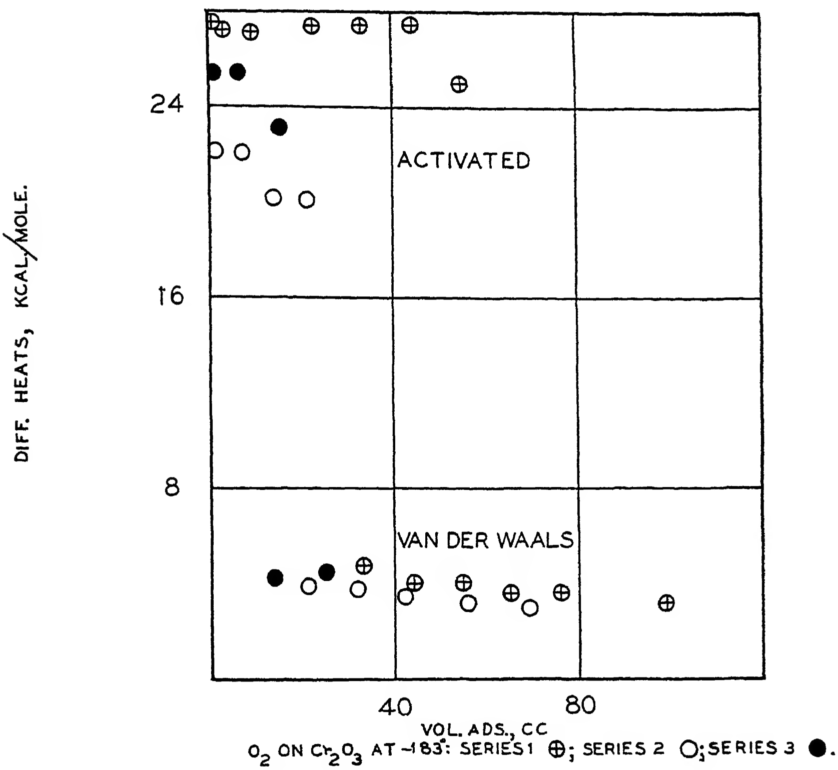


FIG. 10

adsorption frequently has been followed by observing the slow disappearance of the unadsorbed gas in the apparatus. In the present experiments there was practically no residual unadsorbed gas after the first minute, and the slow evolution of heat could be attributed only to a change *on the surface* from the van der Waals to the activated state. Because all the gas disappeared almost instantaneously the calorimetric method offered the only means of detecting the presence of the slow process. It will be interesting to ascertain how general is this occurrence of activated adsorption via an initial van der Waals process. Calori-

metric work at − 183°C now in progress in this laboratory shows that the adsorption of oxygen on iron synthetic ammonia catalyst follows the same complex mechanism.

Acknowledgment

Thanks are due to the American Philosophical Society for a grant-in-aid from the Penrose Fund used for technical assistance in this research.

LIST OF REFERENCES

1 (a) Lamb and Coolidge, *Jour. Am. Chem. Soc.*, **42**, 1146 (1920).
(b) Keyes and Marshall, *ibid.*, **49**, 156 (1927).
(c) Marshall and Bramston-Cook, *ibid.*, **51**, 2019 (1929).
(d) Russell and Gehring, *ibid.*, **57**, 2544 (1935).
2 Dewar, *Proc. Roy. Soc.* (*London*), **74**, 122 (1904).
3 Pearce and McKinley, *J. Phys. Chem.*, **32**, 360 (1928).
4 (a) Taylor and Kistiakowsky, *Zeit. physik. Chem.*, **125**, 341 (1927).
(b) Magnus and Kälberer, *Zeit. f. anorg. Chem.*, **164**, 345 (1927).
(c) Garner and Veal, *Jour. Chem. Soc.*, 1436 (1935).
(d) Beebe and Orfield, *Jour. Am. Chem. Soc.*, **59**, 1627 (1937).
5 (a) Schwab and Brenneke, *Z. physik. Chem.*, 16B, 19 (1932).
(b) Beebe, *Trans. Farad. Soc.*, **28**, 761 (1932).
6 Burwell and Taylor, *Jour. Am. Chem. Soc.*, **58**, 697 (1936).
7 Gould, Bleakney and Taylor, *J. Chem. Phys.*, **2**, 362 (1934).

THE IḤYĀ' 'ULŪM AL-DĪN OF AL-GHAZZĀLI

NABIH AMIN FARIS

Department of Oriental Languages and Literatures,
Princeton University
(Communicated by Cyrus Adler)
(*Read Nov. 19, 1938*)

ABSTRACT

The Iḥyā' 'Ulūm al-Dīn is the *magnum opus* of al-Ghazzāli, the greatest Moslem theologian and one of the most original thinkers of Islam. The book consists of four volumes: the first two treat of the outward forms of worship, while the last two deal with the inward nature of religion. Through it, al-Ghazzāli insured for mysticism a firm and permanent position in the "Church" of Islam, established its orthodoxy, and brought philosophy and philosophical theology within the range of the ordinary man. Indirectly its influence left a deep impress on Christian and Jewish scholasticism. Its study, therefore, is of great importance for a complete understanding of mediæval thought and culture in both the East and the West. The present study is the first step in that direction.

NOTHING which concerns the life of one-sixth of the population of the earth could be dismissed as something merely academic, remote from the pressing problems of this world. On the contrary, it should attract the interest as well as the careful consideration of all thoughtful men. For this reason I consider the study of the *Iḥyā' 'Ulūm al-Dīn* of al-Ghazzāli [1] as useful knowledge, wholly within the scope of this Society. Still more, in that study, without doubt, is the hope for the future of Islam.

The second half of the fifth Moslem century saw a revival in the temporal power of Islam under the strong hand of the newly Islamized Saljūqs. Ṭughril Beg as well as Alp Arslān, Malik Shāh, and their great vizier, the illustrious Niẓām-al-Mulk, rescued a dying state, re-established its power, and gave it a new lease on life. But these potentates did not, and could not, give Islam a religious reformation or a cultural renaissance. The need for such a revival did not escape the notice of the great Niẓām-al-Mulk. He, therefore, founded and endowed [2] the Niẓāmīyah school in Baghdad. This was the first real academy in Islam which made provisions for the physical needs of its students [3]

[1] A.H. 450–505/A.D. 1059–1111. See ibn-Khallikān, *Wafayāt al-A'yān* (Cairo, 1299), vol. II, pp. 246–7; al-Subki, *Tabaqāt al-Shāfi'īyah al-Kubra* (Cairo, 1324), vol. IV, pp. 101–82; D. B. Macdonald, "The Life of al-Ghazzālī" in *Journal of the American Oriental Society,* vol. XX (1899), pp. 71–132.

[2] A.H. 457–9/A.D. 1065–7.

[3] See al-Suyūṭi, *Ḥusn al-Muḥāḍarah fi Akhbār Miṣr w-al-Qāhirah* (Cairo, 1321), vol. II, pp. 156–7.

and became a model for later institutions of higher learning. He called to it the leading scholars of the time. Among those scholars was al-Ghazzāli, the author of the *Iḥyā' 'Ulūm al-Dīn.*

It did not take al-Ghazzāli long to conclude that without a corresponding ethical and moral reawakening, the temporal gains which the Saljūqs had achieved would not endure. He even declared that such a culture was purely materialistic, and, therefore, it should not endure. He diagnosed the disease and prescribed a remedy, nay, the remedy. This remedy was education. Yet it was an education unlike that which the best minds of his time advocated. Theirs was limited to the imparting of knowledge for utilitarian purposes. His emphasized the necessity of stimulating the moral consciousness of the individual, thus becoming the first teacher in Islam to bring the problem of education into organic relation with a profound ethical system. In the midst of a period which, in many ways, was not unlike our own, he sought a unifying principle for life; and unlike some of the savants of our time, he was not afraid of the word religion. Through it, life which seems disunited, useless, and unhappy, becomes united, useful, and happy. And to that end he wrote his *magnum opus*, the *Iḥyā' 'Ulūm al-Dīn* (The Revival of the Sciences of Religion).

No critical edition of this epoch-making book has yet been made.[1] Its value has, therefore, been greatly limited. The work which I have done on the *Iḥyā'* and which was made possible by the generosity of the American Philosophical Society has laid down the foundation for further and more detailed study of al-Ghazzāli's accomplishments and influence upon the Moslems in particular and the Mediæval West in general.

Naturally establishing the text had to come first. This was based on the collation of four available texts: a fifteenth century manuscript copy of the work, belonging to Mr. Robert Garrett of Baltimore and deposited in the Princeton University Library, a recordak copy of a seventeenth century manuscript in the British Museum, the Cairo text which reproduces the manuscript copy of the Khedivial Library, and the text contained in the commentary of al-Sayyid al-Murtaḍa.[2] After the Arabic text of the first 'quarter' had been established, an English translation was prepared. In both the text and the translation biographical, historical, legal, theological, and linguistic notes were added. A special attempt has been made to identify every name of person or

[1] D. B. Macdonald has translated Kitāb al-Samā', Book 8 of Rub' al-Mu'āmalāt, in *Journal of the Royal Asiatic Society* (1901–2); Edwin E. Calverley has also published a translation of Kitāb al-Ṣalāh, Book 4 of Rub' al-'Ibādāt, under the title *Worship in Islam* (Madras, 1925). A great part of the *Iḥyā'* has been analyzed by Miguel Asin in his *Algazel.* A German translation is being prepared by H. Bauer.

[2] *Itḥāf al-Sādah al-Muttaqīn bi-Sharḥ Asrār Iḥyā' 'Ulūm al-Dīn*, in 10 volumes (Cairo, 1311)

place, check every tradition and cite its source, and determine Koranic references. It is hoped to have the Arabic text and the English translation published jointly in one book, and thus, for the first time, place at the disposal of the learned world a standard edition which would serve as the basis for further research and investigation into the life and influence of al-Ghazzāli.

The work itself is divided into four parts. This is no mere accident. Desiring to insure for his ideas the widest circulation possible, al-Ghazzāli modelled his *Iḥyā'*, only in form, after the most popular books of the day. These dealt with jurisprudence, and were always divided into four parts, one for each of the component parts of the discipline, namely the Koran, the *sunnah* (the usage of the Prophet as recorded in tradition), catholic consent, and analogy. Al-Ghazzāli was not the first Moslem writer to employ this device. He himself alludes to the *Taqwīm al-Abdān* (Tables of Physiology) of ibn-Jazlah [1] (d. A.H. 493/A.D. 1100), which, like the earlier *Taqwīm al-Ṣiḥḥah* (Tables of Health) of ibn-Buṭlān [2] (d. A.H. 455/A.D. 1063), was a medical work modelled, for the purpose of gaining a wider audience, after the then very popular astronomical tables. Each of these four parts of the *Iḥyā'* was called *rub'* (quarter). The first deals with the Acts of Worship (*'ibādāt*), the second treats of the Usages of Life (*mu'āmalāt*), the third discusses the Destructive Matters of Life (*muhlikāt*), and the fourth expounds the Saving Matters of Life (*munajjiyāt*). The first two deal with the outward forms of worship, while the last two treat of the inner nature of religion.

Each of these four 'quarters' contains ten books (sing. *kitāb*). The quarter on the Acts of Worship comprises the following:

1. The Book of Knowledge
2. The Articles of Faith
3. The Mysteries of Purity
4. The Mysteries of Prayer
5. The Mysteries of Alms
6. The Mysteries of Fasting
7. The Mysteries of the Pilgrimage
8. The Rules of Reading the Koran
9. On Invocations and Supplications
10. On the Office of Portions

[1] See ibn-Khallikān, vol. III, pp. 255–6; ibn-abi-Uṣaybi'ah, *'Uyūn al-Anbā' fi Ṭabaqāt al-Aṭibbā'* (Cairo, 1299), vol. I, p. 255.

[2] See ibn-abi-Uṣaybi'ah, vol. I, pp. 241–3.

The quarter on the Usages of Life comprises the following:

1. The Ethics of Eating
2. The Ethics of Marriage
3. The Ethics of Earning a Livelihood
4. On the Lawful and the Unlawful
5. The Ethics of Companionship and Fellowship with the Various Types of Men
6. On Seclusion
7. The Ethics of Travel
8. On Audition and Grief
9. On Enjoining Good and Forbidding Evil
10. The Ethics of Living as Exemplified in the Virtues of the Prophet

The quarter on the Destructive Matters of Life comprises the following:

1. On the Wonders of the Heart
2. On the Discipline of the Soul
3. On the Curse of the Two Appetites—the Appetite of the Stomach and the Appetite of Sex
4. The Curse of the Tongue
5. The Curse of Anger, Rancour, and Envy
6. The Evil of the World
7. The Evil of Wealth and Avarice
8. The Evil of Pomp and Hypocrisy
9. The Evil of Pride and Conceit
10. The Evil of Vanity

The quarter on the Saving Matters of Life comprises the following:

1. On Repentance
2. On Patience and Gratitude
3. On Fear and Hope
4. On Poverty and Asceticism
5. On Divine Unity and Dependence
6. On Love, Longing, Intimacy, and Contentment
7. On Intentions, Truthfulness, and Sincerity
8. On Self-Examination and Self-Accounting
9. On Meditation
10. On Death.

In these forty books which make up the four quarters of the *Iḥyā'*, al-Ghazzāli has preserved the summation of Mediæval Moslem thought.

For this reason the *Iḥyā'* occupies a unique position throughout the Moslem world. This position is best described in the words of Ḥājji Khalīfah, the foremost Turkish writer of the seventeenth century, who said, "Should all other Moslem writings be destroyed, the *Iḥyā'*, if spared, would make up for all the loss." [1]

Through the *Iḥyā'*, al-Ghazzāli grafted mysticism, which had hitherto been deemed unorthodox, to Islam, and established its orthodoxy. His mysticism vitalized the law by making personal religion and individual experience a part of Islam. His orthodoxy safeguarded the faith against unbridled emotionalism. This won for him the honorific title of The Authority of Islam (*Ḥujjat al-Islām*). Through it he led the Moslems back from scholastic labours upon theological dogma and minutiæ to a living contact with the Word. Through it he brought philosophy, which he regarded merely as *thinking*, and philosophical theology within the range of the ordinary man.

But the influence of al-Ghazzāli extended beyond the walls of Islam. His works were partly translated into Latin before the middle of the twelfth century,[2] and have left a marked influence upon Jewish and Christian scholasticism. His influence upon Thomas Aquinas (d. 1274) came through the *Pugio Fidei* of Raymundus Martinus (XIIIth cent.). The extent of this influence still has to be determined. This study is but the first step in that direction.

[1] Ḥājji Khalīfah, *Kashf al-Ẓunūn fi Asmā' al-Kutub w-al-Funūn*, ed. G. Flügel (Leipzig and London, 1835–58), vol. I, p. 180.

[2] See *The Legacy of Islam*, ed. Thomas Arnold and Alfred Guillaume (Oxford, 1931), pp. 269–74.

VASCULARITY IN THE BRAINS OF TAILED AMPHIBIANS

I. *Ambystoma tigrinum* (Green)[1]

E. HORNE CRAIGIE

Department of Biology, University of Toronto

(Communicated by Edwin G. Conklin)

ABSTRACT

The caliber of the capillaries in most parts of the brain of *Ambystoma tigrinum* is somewhat smaller than in the leopard frog. The widest vessels were found in the hypoglossal nucleus, the narrowest in the neuropil of the corpus striatum.

Vascularization is poorer than in the frog in most regions. The nucleus VIII is richer than the vestibular and poorer than the cochlear nucleus of the frog, and is the richest center found in *Ambystoma*.

The forebrain neuropil is richer than in the frog, where it was very poor. It is richer also than the corresponding cellular area and than any other part studied in the cerebral hemisphere, which in general is poorly vascularized. The molecular layer of the cerebellum, also, is richer than the granular layer.

Cerebral capillaries are narrower and in most regions less numerous than those in the pars distalis of the hypophysis.

It has been known since the time of Schöbl ('82) that tailed amphibians differ fundamentally from the tailless members of that order, as well as from the majority of other vertebrates in respect of the arrangement of the capillaries in the tissues of the central nervous system. Roofe ('35), however, pointed out that *Ambystoma tigrinum* has a network of capillaries like that in the *Salientia* and the present author ('38 a) reported that *Ambystoma* differed from all other genera hitherto examined in that it has both a capillary network, though a very simple one, and a system of independent capillary loops such as occurs in other Caudata. Speculations as to the possible physiological significance of such slender loops were recently published (Craigie, '38 b).

In view of the architectural differences just cited as well as of the less highly specialized character of the brain in Caudata as compared with *Salientia*, it seems natural to wonder whether there are significant differences in the richness of the capillary supply in these groups. A report of a quantitative study of the vascularity of the brain of the frog is in press ('38 d). Since *Ambystoma* differs from other Caudata in vascular arrangement, as indicated above, measurements in a member of that genus are here submitted, while a similar study of *Necturus*, which has only capillary loops, is to follow.

[1] Study aided by a grant from the Penrose Fund of the American Philosophical Society.

Material

The same ten brains of *Ambystoma tigrinum* used in the study of the vascular arrangement cited above ('38 a) and in that of the hypophysis ('38 c) were employed in the present investigation. They were derived from two males and two females obtained from a Chicago dealer and from four males and two females captured in Manitoba. The blood vessels were injected with carmine gelatine. The measurements were made in serial sections 20 microns thick, while thick sections were cut at intervals, which helped to serve as a check on the completeness of the injection. It may be pointed out, however, that examination of thick sections is a less useful check on completeness of injection when isolated capillary loops are in question than when there is only a continuous network. The microscope and micrometer were the same employed in all the writer's investigations of this kind.

The injection was not perfect in all ten specimens and in a few cases, particularly in the cerebellum, measurements were omitted or discarded, which fact explains the gaps in Table 1. Wherever one specimen gave a markedly lower result than the rest it was carefully checked. In most cases those specimens which did not appear quite satisfactorily injected gave values within the range of variation of the others, sometimes rather high, and such values have been retained. It is hoped that with these safeguards the general results are reasonably trustworthy. It is noticed, however, that the individual variation and, hence, the probable errors are rather large.

The identification of the regions studied is based upon the papers of Herrick ('14, '27, '30).

Caliber of the Capillaries

The caliber of the capillaries was measured in seven regions in each of five specimens. In most cases the value obtained in each specimen is the average of fifty measurements of single vessels, though in a few regions capillaries are so few that thirty were averaged.

In the former study ('38 a) no significant difference was found between the average diameter of the capillaries forming part of the continuous network and those forming independent loops in the same region, despite the fact that one limb of a loop was usually wider than the other. Hence, in the present research, no attempt was made to consider the two types of capillaries separately.

The results of the measurements are presented in the first column of Table 3. The diameter in the molecular layer of the cerebellum and

that in the primordium hippocampi are identical with those in the leopard frog while the others are somewhat less than in the latter. As in the frog, the widest vessels are found in the hypoglossal nucleus, the narrowest in the neuropil of the corpus striatum. Other relations noted in the frog, however, were not found in *Ambystoma*.

The diameters are mostly similar to those in the dogfish, though the average is a little less, but they are larger than those in the ratfish.

In no part of the brain is the caliber of the capillaries as large as in the epithelial lobes of the hypophysis in the same animals ('38 c). (The pars nervosa of the hypophysis lacks vessels within its substance.)

Lengths of the Capillaries

The total length of the capillaries occurring in a unit volume of tissue was measured in the same regions which had been studied in the

TABLE 1

Total Length in Microns of the Capillaries in $\frac{1}{2} \times 189^2 \times 200$ c.μ of Tissue in the Brain Stem and Cerebellum of the Tiger Salamander (not Corrected for Shrinkage)

	Ambystoma										Average	% P.E. of Average	Corrected Average Length in 100^3 c.μ of Fresh Tissue
	5 ♀	6 ♂	13 ♂	14 ♀	15 ♂	16 ♂	17 ♂	19 ♂	20 ♂	21 ♀			
Fasc long medialis	243	307	336	320	288	295	221	205	201	297	271	3.9	61
Nuc. motorius dors. X	479	597	408	383	497	562	429	357	572	442	473	3.8	106
Nucleus XII	618	308	315	569	313	404	512	432	436	572	448	5.5	100
Nuc. motorius VII	415	623	489	391	518	590	564	354	425	584	495	4.1	111
Nuc. motorius V	769	443	462	452	604	559	619	450	482	543	538	4.1	121
Nuc. fasc solitarii	409	267	168	401	207	241	216	196	321	423	285	7.2	64
Nuc spinalis V	377	456	375	333	406	467	345	475	469	379	408	2.8	91
Cerebellum: molecular layer	684	721	—	564	—	—	525	360	330	767	565	7.8	127
Nuc. sens principalis V	594	820	650	768	792	546	772	526	752	804	702	3.4	157
Cerebellum: granular layer	541	481	423	350	194	—	205	338	205	269	334	8 6	75
Eminentia cerebelli ventr.	536	571	485	303	303	—	340	359	321	308	392	6.2	88
Nucleus VIII	774	978	752	676	724	986	952	766	781	1019	841	3.2	188

TABLE 2

Total Length in Microns of the Capillaries in $\frac{1}{2} \times 189^2 \times 200$ c.μ of Tissue in the Forebrain of the Tiger Salamander (not Corrected for Shrinkage)

	Ambystoma										Average	% P.E of Average	Corrected Average Length in 100^3 c.μ of Fresh Tissue
	5	6	13	14	15	16	17	19	20	21			
Primordium hippocampi	559	356	595	331	412	349	551	354	549	404	446	5.0	100
Primordium pallii dorsalis	297	268	260	249	155	250	270	170	208	168	230	4.7	52
Primordium piriforme	179	180	171	148	194	161	124	122	138	189	161	3.5	36
Neuropil of p. piriforme	567	401	403	475	559	688	563	615	477	590	534	3.7	120
Nucleus medialis septi	537	236	426	397	290	372	428	420	457	459	402	4.6	90
Corpus striatum, pars ventralis	175	70	96	147	213	98	182	188	221	112	150	7.6	34
Neuropil of corpus striatum, pars ventralis	732	867	600	531	662	774	685	544	608	453	628	3.4	141
Nucleus dorsolateralis amygdalae	277	288	111	235	210	255	127	163	205	145	202	6.6	45

frog, with the exception of the cochlear and superior olivary nuclei which are not differentiated in *Ambystoma*. The results are presented in Tables 1 and 2. The correction for shrinkage (coefficient 0.8) used in obtaining the figures in the last column of each of these tables is only rough, being based upon shrinkage actually measured in the whole brain of the frog.

It may be questioned whether capillaries disposed as simple loops are strictly comparable from a quantitative functional point of view with those forming a spongy network but in the present investigation this distinction was regarded as unimportant.

The averages show the vascularization to be considerably poorer than in the leopard frog ('38 d) in all except three regions, namely, the nucleus of the eighth nerve and the two areas of neuropil in the cerebral hemisphere. The nucleus VIII is markedly richer than the vestibular nucleus of the frog but much poorer than the cochlear nucleus of that animal, a condition which may perhaps be attributable to its containing the undifferentiated representatives of both these centers. Like the dorsal cochlear nucleus in other animals studied, it is distinctly the richest center examined.

The difference in vascular richness of the forebrain neuropil is very striking. In the frog the two areas of this formation were found to be the poorest parts of the gray matter and not to differ from the medial longitudinal bundle to a statistically significant extent. In *Ambystoma*, on the other hand, the two neuropil areas are markedly richer than any other part of the cerebral hemisphere studied and rank well up among the gray regions of the hindbrain. This remarkable difference in the relative vascularity of the neuropil was discussed at some length in the paper on the brain of the frog, the richness found in *Ambystoma* being in agreement with what had already been observed by Herrick and by the present writer in *Necturus*. The poverty of the gray matter as compared with the overlying white in respect of capillaries is clearly visible in the accompanying photograph (Fig. 1), which shows also the more uniform distribution of the capillaries through the thickness of the medial wall of the hemisphere.

Apart from the neuropil, the cerebral hemisphere is poorly vascularized, the richest areas examined being the primordium hippocampi and the nucleus medialis septi which rank with the poorer regions in the hindbrain, while the other two pallial primordia, the corpus striatum, and the nucleus dorsolateralis amygdalæ are all poorer than the medial longitudinal bundle. The order of richness in these forebrain centers is essentially the same as in the frog.

As in the frog, the granular layer of the cerebellum is a poorly vascular region. In both amphibians the molecular layer of the cerebellum is richer than the granular, particularly in *Ambystoma*, a relation the reverse of that in either the elasmobranchs or the mammals but corresponding with the richness of the neuropil in the cerebral hemisphere of the salamander. Despite the rather large probable errors for both regions, the difference is fully significant statistically.

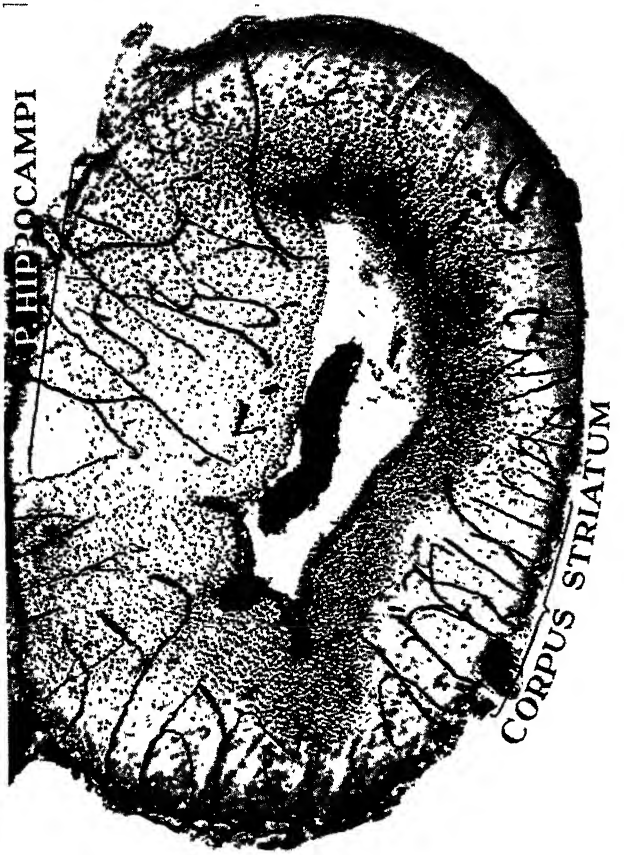


Fig. 1. Photomicrograph of transverse section, 150 μ thick, of right cerebral hemisphere of *Ambystoma tigrinum*. Mag. $\times$ 50. The small number of capillaries in the cellular regions of the corpus striatum and of the primordium piriforme in contrast with the overlying white matter and neuropil is evident, as is the more uniform distribution of vessels through the thickness of the wall in the region of the primordium hippocampi.

The vascularity of most parts of the brain is a good deal less than that of the pars distalis of the hypophysis ('38 c).

The volumes of the capillaries and the areas of their walls in fresh tissue in the seven regions where the diameters were measured are

presented in the second and third columns of Table 3. The correction for shrinkage used in calculating these values was based upon shrinkage observed in the frog (viz. linear shrinkage of 10.7 per cent). The volumes and areas are all smaller than in the frog except in the neuropil of the corpus striatum, the vascular richness of which is reflected in these dimensions as in the length of the capillaries, despite their small caliber.

TABLE 3

VARIOUS DIMENSIONS OF THE CAPILLARIES IN PARTS OF THE BRAIN OF THE TIGER SALAMANDER

	Average Diameter in Microns, not Corrected	Volume in cu μ of Capillaries in 100^3 cu μ Fresh Tissue	Area in sq. μ of Cap. Walls in 100^3 cu μ Fresh Tissue	Area in sq mm. to Which 1 cu. mm. of Blood is Exposed
Fasc. long. medialis	5.4	1724	1150	667
Nucleus XII	5.7	3216	2009	625
Cerebellum, molecular layer	4.9	3016	2195	727
Cerebellum, granular layer	4.7	1654	1248	755
Corpus striatum, pars ventralis	5.3	929	630	678
Neuropil of corpus striatum	4.6	2879	2259	784
Primordium hippocampi	5.6	3117	1979	635
Average	5.2			696

In the last column of Table 3 is shown the area of capillary wall over which one cubic millimeter of blood is spread in each of the same regions. Reflecting the smaller caliber of the vessels, these areas are all larger than in the frog except that in the molecular layer of the cerebellum and that in the primordium hippocampi.

SUMMARY

The caliber of the capillaries in most parts of the brain of *Ambystoma tigrinum* is somewhat smaller than in the leopard frog. In the molecular layer of the cerebellum and in the primordium hippocampi the caliber is the same in these two species. The widest vessels were found in the hypoglossal nucleus, the narrowest in the neuropil of the corpus striatum.

Vascularization is poorer than in the frog in most regions. The nucleus VIII is richer than the vestibular and poorer than the cochlear nucleus of the frog, and is the richest center found in *Ambystoma.*

The forebrain neuropil is richer than in the frog, where it was very poor. It is richer also than the corresponding cellular area and than any other part studied in the cerebral hemisphere, which in general is poorly vascularized. The molecular layer of the cerebellum also is richer than the granular layer.

Cerebral capillaries are narrower and in most regions less numerous than those in the pars distalis of the hypophysis.

LITERATURE CITED

CRAIGIE, E. HORNE. 1938 a. The Blood Vessels of the Brain Substance in some Amphibians. *Proc. Amer. Phil. Soc.*, **78**, 615–649.

——. 1938 b. The Blood Vessels in the Central Nervous System of the Kangaroo. *Science*, N.S., **88**, 359–360.

——. 1938 c. The vascularization of the Hypophysis in Tailed Amphibians. *Trans. Roy. Soc. Can.*, Sec. V, Ser. III, **32**, 43–50.

——. 1938 d. Vascularity in the Brain of the Frog (*Rana pipiens*). *J. Comp. Neur.*, **69**, 453–470.

HERRICK, C. JUDSON. 1914. The Medulla Oblongata of Larval *Amblystoma*. *J. Comp. Neur.*, **24**, 343–427.

——. 1927. The Amphibian Forebrain. IV. The Cerebral Hemisphere of *Amblystoma*. *J. Comp. Neur.*, **43**, 231–325.

——. 1930. The Medulla Oblongata of *Necturus*. *J. Comp. Neur.*, **50**, 1–96.

ROOFE, P. 1935. The Endocranial Blood Vessels of *Amblystoma tigrinum*. *J. Comp. Neur.*, **61**, 257–293.

SCHÖBL, J. 1882. Ueber die Blutgefässe des cerebrospinalen Nervensystems der Urodelen. *Arch. f. mikr. Anat.*, Bd. 20, S. 87–92.

LAND MOLLUSKS FROM CORN ISLAND, NICARAGUA [1]

HORACE G. RICHARDS

New Jersey State Museum, Trenton, New Jersey

(Communicated by Edwin G. Conklin)

ABSTRACT

Great Corn Island lies in the Caribbean Sea about 40 miles east of Bluefields, Nicaragua. The island is about two and a half miles long by two miles wide and rises at its highest point to 370 feet above the sea. It is entirely composed of igneous rock of late Tertiary or early Pleistocene age. There is no Pleistocene limestone such as is found on certain other islands off the Central American coast. The molluscan fauna is exceedingly poor, consisting almost entirely of "drift" species common throughout all of Tropical America and of species described from the mainland of Central America. There was one new species. The absence of West Indian forms is noteworthy. There are several possible explanations for this poor fauna. In the first place the absence of limestone is frequently correlated with a poor fauna, particularly in this part of the world. Furthermore, much of the island is under cultivation and this may have caused the depletion of the original fauna.

GREAT Corn Island lies in the Caribbean Sea at about 12° 7′ north latitude and 83° west longitude, and about 40 miles east-north-east of Bluefields, Nicaragua. The island is only about two and a half miles in length by two miles in width (Fig. 1).[2] It is relatively low, the highest point, Mount Pleasant, being only 370 feet above the sea. The next highest point, Quin Bluff, near the south end of the island reaches a height of about 100 feet. With the exception of this bluff, the shore line is low and sandy. A large part of the island is cultivated with coconuts and bananas. Mangrove swamps frequently extend back from the shore line and swampy depressions are frequent throughout the entire island.

Little Corn Island lies about 7 miles north-north-east of Great Corn Island and is much smaller, being only about a mile and a half long by half a mile in width. It rises only a little more than 100 feet. The term Corn Island without the adjective always refers to the larger island.

GEOLOGY

Great Corn Island is composed of "young eruptive rocks" (Sapper, 1937, p. 113, Pl. 8).[3] It is probable that these rocks date from the Late

[1] With the support of a grant from the Penrose Fund of The American Philosophical Society.

[2] The accompanying map (Fig. 1) was drawn by my companion, Allen L. Midyette, Jr., and is based in part upon the British Hydrographic survey of 1836. (See U. S. Hydrographic Chart No. 1517.)

[3] Some rock samples from Quin Bluff were identified by Dr. H. H. Hess as "badly weathered basalts containing microphenocrysts of labradorite and pleochroic hypersthene."

Tertiary or Early Pleistocene. Similar rocks outcrop on the Nicaraguan mainland west and northwest of Bluefields (Sapper, 1931, Pl. 8). These rocks are best exposed on the island at Quin Bluff (Fig. 2).

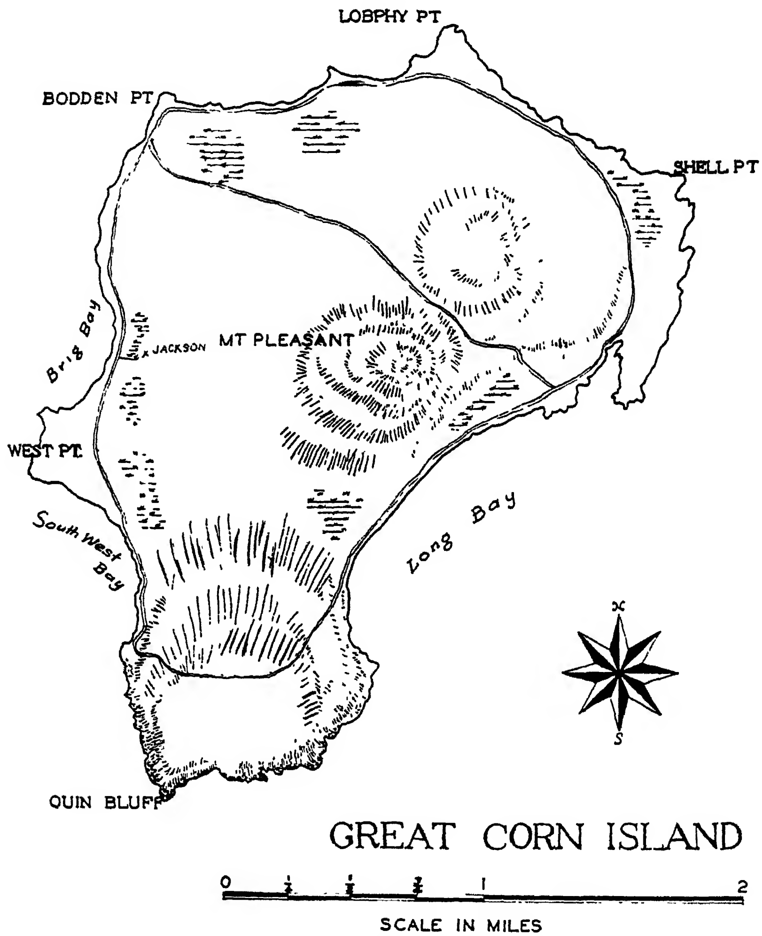


FIG. 1. Map of Great Corn Island, Nicaragua.

Contrary to conditions on many Caribbean Islands, there is no limestone on Corn Island. Many islands in this vicinity contain limestone deposits laid down during the last high sea level stage of the Pleistocene. Cozumel was entirely of limestone and contained Pleistocene marine fossils up to elevation 25 feet above the sea (Richards, 1937). Roatan Island, Honduras, although largely of metamorphic rock, had a Pleistocene limestone cap at its western end which contained fossil coral (Richards, 1938).

The absence of a Pleistocene limestone cap on Corn Island is difficult to explain. It is possible, although not probable, that the island has risen out of the sea since Pleistocene time. It is more probable that Pleistocene limestone was formed and has since eroded. It is equally possible that the island was higher than at present during the Pleistocene and that any limestone laid down during a high sea level stage has since sunk beneath the sea.

FIG. 2. Shore line of Corn Island near Quin Bluff, looking toward Mount Pleasant.

PRESENT WORK

The only naturalists known to have visited the Corn Islands are James L. Peters and Edward Bangs from the Museum of Comparative Zoology of Cambridge, Massachusetts, who collected vertebrates from there during the winter of 1927–28. The fauna was found to be exceedingly poor and to contain few if any species of special interest (Peters et al., 1929).

Since no mollusks had ever been recorded from the islands, it seemed desirable to obtain a collection. Therefore a visit was paid to Great Corn Island in August, 1938. Accompanied by Allen L. Midyette, Jr., I sailed from New Orleans for Bluefields, Nicaragua, on the United Fruit Company's freighter "Sinaloa." At Bluefields we chartered a small boat the "Frank Otis" to take us to Corn Island.

The fauna was found to be very poor, both in species and in individuals. This was probably largely due to the lack of limestone on the Island. Another reason for the paucity of the fauna is the fact that so much of the island is cultivated, resulting in the probable depletion of the mollusks. In fact, the majority of the species are characteristic of cultivated areas and are of widespread tropical distribution. The best locality for collecting mollusks was near Quin Bluff, but not even here were the species abundant.

It is to be regretted that circumstances did not permit a visit to Little Corn Island; however, it was necessary to return to Bluefields with the "Frank Otis" at the completion of our collecting on Great Corn Island.

Previous Literature on Nicaragua Mollusks

As mentioned above, there are no previous published records on the molluscan fauna of the Corn Islands. The most complete works on the land and freshwater mollusks of the Central American region are those of Fisher and Crosse (1870–1900) and Von Martens (1890–1901). The only paper dealing specifically with the fauna of Nicaragua is that of Tate (1870). Other papers on the Central American fauna were listed by the writer (Richards, 1938) in a discussion of the mollusks of Roatan Island, Honduras.

Land Mollusks from Corn Island

In addition to the material collected on the present trip, I have had the opportunity, through the kindness of Dr. William J. Clench of the Museum of Comparative Zoology, Harvard University, of examining the few mollusks obtained by Peters in 1927. Only one species was represented in their collections, *Helicina rostrata* Morelet.

Oleacinidæ

Salassiella cornii new species (Fig. 3)

Shell minute, subfusiform, thin, smooth with small irregular striæ; dead white, chalky (weathered); spire conic, ogival but with blunt apex; suture weakly impressed but more so than in *S. modesta;* whorls 4, slightly convex, the last whorl slightly longer than the spire; columella obliquely truncate, weakly concave; aperture slightly wider than in *S. modesta.*

Type: Length 4.5 mm. (Acad. Nat. Sci. Phila. 173926)

Type locality: Near Quin Bluff, Corn Island, Nicaragua.

Closely related to *S. modesta* (Pfeiffer) from Mirador and Orizaba in the State of Vera Cruz, Mexico, but distinguished from it by having

its sides more nearly parallel, by its more impressed suture, by its less conic spire, by its obliquely instead of abruptly truncate columella and by its slightly wider aperture.

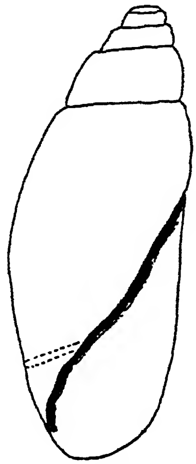


FIG. 3. *Salassiella cornii* Richards; Type specimen × about 7.

S. perpusilla (Pfeiffer) from the States of Vera Cruz and Nuevo Leon, Mexico, and Guatemala, is another related species, but is distinguished by being more convex below the suture.

This is the first record of the genus *Salassiella* from Central America south of Guatemala.

Sagdidæ

Thysanophora cæcoides (Tate)

Near Quin Bluff.

Widely distributed throughout Central America and Yucatan; originally described from Chontales Forest, Nicaragua (Tate, 1870); Cozumel.

Thysanophora plagioptycha (Shuttleworth)

Near Quin Bluff.

Widespread throughout Central America and the West Indies; Cozumel; Swan Island; Venezuela. Frequently associated with cultivated ground.

Bulimulidæ

Oxystyla princeps (Sowerby)

South end Great Corn Island.

A common arboreal species from tropical Mexico and Central America; Cozumel.

Subulinidæ

Lamellaxis micra (d'Orbigny)

Opeas micra (d'Orbigny)

South end Great Corn Island; near Quin Bluff.

Widespread throughout all Tropical America, probably largely carried by commerce; Mexico; Cozumel; Guatemala; Roatan, Honduras; Polvon and Ometepe, Nicaragua; Costa Rica; West Indies.

Lamellaxis mexicanus (Pfeiffer)

Leptinaria mexicana (Pfeiffer)

South end Great Corn Island; near Quin Bluff.

Known from Mexico; Guatemala; Roatan Island, Honduras; and Ometepe Island, Nicaragua.

Subulina octona Bruguiere

Near Quin Bluff.

Widespread through all the tropics.

Succineidæ

Succinea recisa Morelet

Near Quin Bluff.

Reported from Guatemala; Roatan, Honduras; central and southwest Nicaragua; Costa Rica; Panama.

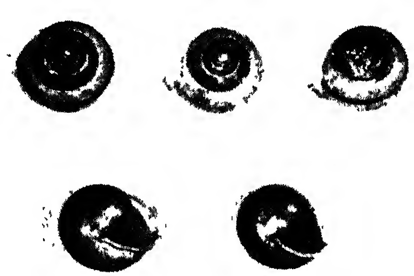


FIG. 4. *Helicina rostrata* Morelet, × about 2.

Helicinidæ

Helicina rostrata Morelet (Fig. 4)

Corn Island (Peters).

This species is known from Vera Paz and Coban in Guatemala and from San Diego, Acoyapa and La Luz Mine (Suina) in Nicaragua. It is

very close to, if not actually identical with, *H. denticulata* Pfeiffer which has been reported from Honduras and East Nicaragua.

Helicina fragilis Morelet

Near Quin Bluff.

Mexico (States of Vera Cruz and Guerrero); Peten, Guatemala.

Lucidella lirata (Pfeiffer)

South end Great Corn Island; near Quin Bluff.

Common on the mainland from southern Mexico to Panama and Venezuela.

Zoogeography

The water between Corn Island and the mainland is relatively shallow and does not exceed 20 fathoms in depth. Therefore, even a moderate lowering of sea level during the Pleistocene would have caused the two to be connected. This probably accounts for the close relationship between the Corn Island fauna and that of the mainland. The one new species (*Salassiella cornii*) probably also occurs on the Nicaraguan mainland, but has not yet been recorded due to insufficient collecting.

Since the island is of Late Tertiary or Early Pleistocene age, it could not have formed part of the ancient land connection which, according to Schuchert (1936, pp. 107–108), joined Jamaica and Haiti with Nicaragua and Honduras during early Tertiary time. It was suggested in a previous paper (Richards, 1938, p. 177) that this land bridge might have permitted the introduction of certain species into Roatan Island, Honduras and the adjacent mainland.

The poor fauna of Corn Island can probably best be explained by the absence of limestone which would be the most suitable habitat for mollusks, and because of the probable recent connection between the Island and the mainland. The fauna of the east coast of Nicaragua is notably poor.

Bibliography

Fisher, P. and Crosse, H. 1870–1900. Etude sur les mollusques terrestres et fluviatiles du Mexique et du Guatemala. *Mission scientifique au Mexique et dans l'Amerique Centrale. Recherches Zoologiques*, Pt. 7, 1 and 2.

Peters, James; Allen, Grover; Barbour, Thomas; and Loveridge, Arthur. 1929. Vertebrates from the Corn Islands. *Bull. Mus. Comp. Zool.*, **69**, 127–146.

Richards, H. G. 1937. Land and Freshwater Mollusks from the Island of Cozumel, Mexico, and their Bearing on the Geological History of the Region. *Proc. Amer. Phil. Soc.*, **77**, 248–262.

——. 1938. Land Mollusks from the Island of Roatan, Honduras. *Proc. Amer. Phil. Soc.*, **79**, 167–178.

SAPPER, KARL. 1937. Mittelamerika. *Handbuch der Regionalen Geologie*, 29 heft, Band 8, 4a. Carl Winter's Universitätsbuchhandlung, Heidelberg.

SCHUCHERT, CHARLES. 1936. Historical Geology of the Antillean-Caribbean Region. John Wiley and Sons, New York.

TATE, RALPH. 1870. On the Land and Freshwater Mollusca of Nicaragua. *Amer. Jour. Conch.*, 5, 151–164.

VON MARTENS, E. 1890–1901. Land and Freshwater Mollusca. *Biologia Centrali-Americana*, Zoologia, 9.

STARCH DISSOLUTION AND AMYLOLYTIC ACTIVITY IN LEAVES

H. A. SPOEHR AND HAROLD W. MILNER

Carnegie Institution of Washington, Division of Plant Biology,
Stanford University, California

ABSTRACT

More precise knowledge of the conditions affecting certain enzymes in the living leaf is of importance for an understanding of the processes resulting in the accumulation and disappearance of the products of photosynthesis. For this purpose amylase deserves special consideration, because starch is the first visible product of photosynthesis and also serves as transitory reserve food material. Certain leaves are capable of undergoing starch dissolution in an atmosphere free of oxygen, although the rate of this activity under these conditions is lower than that in normal air. Starch dissolution does not occur in leaves which have been killed by means which do not destroy the enzyme. Higher concentrations of carbon dioxide exert a distinctly inhibiting effect on the rate of starch dissolution of leaves. Comparative determinations of the amylolytic activity of a number of different species of leaves have been made. These show great variation, and also considerable range in the pH of maximum activity of different species. Both starch dissolution and amylolytic activity are decidedly accelerated with the loss of water from the leaves. Also, with decreased temperatures some leaves exhibit an increased amylolytic activity and, in some cases, an increased rate of starch dissolution.

INTRODUCTION

IN the leaves of many plants starch accumulates during periods of illumination. This starch disappears in darkness or when the illumination is insufficient to carry forward the photosynthetic process at a relatively high rate. This process of starch dissolution is brought about by the hydrolytic activity of the enzyme amylase, the presence of which can easily be demonstrated in starch-containing leaves. Amylase has also been isolated from many tissues, both plant and animal, its properties and the conditions influencing its activity have been extensively studied *in vitro*, and are fairly well understood. Although much can be learned concerning enzymes from a study of the action of isolated enzyme preparations *in vitro*, it has become manifest that it is desirable to have more information concerning the activity of the enzymes within the living cell in order to interpret more precisely the physiological phenomena related to enzyme activity. The living cell presents a system and medium which cannot be reproduced by *in vitro* experiments. For example, the effect produced on a leaf by changes in the environment may be brought about through alterations in the activity of an enzyme; yet these same environmental conditions may produce entirely different effects on the isolated enzyme preparation as compared

with those produced on the enzyme acting within the organized cell of the leaf. Similarly, agencies which are known to have no direct effect on an isolated enzyme, may produce a profound effect on the activity of this enzyme when they are applied to the cells containing this enzyme. This does not mean that the nature of the enzyme within and without the cell need differ in any fundamental respect, but rather that within the cell there exist complex and reciprocal relationships which present a very different medium from that of any prepared, *in vitro* system.

The purpose of these investigations was to determine the manner in which the enzymatic activity of living chlorophyllous leaves is influenced by certain environmental factors. In this, special interest was devoted to such conditions as may affect the life or protoplasmatic activity of the cells of the leaf, but which have no direct inactivating influence on the enzymes. In other words, the activity of amylase has been used as an indicator of the manner in which various conditions influence enzyme action in leaves. After considerable preliminary investigation amylase was considered in preference to other enzymes because of its relation to the carbohydrate economy of the leaf, because its properties are fairly well understood, and because of the fact that reliable methods have been worked out for the quantitative determination of even small amounts of starch in leaves.

The development of our knowledge of the presence and function of amylase in leaves has been admirably reviewed by Brown and Morris (1893) in their classical paper on this subject. For many years relatively little attention has been paid to this subject, particularly in regard to the activity of the enzyme in living leaves or, more precisely, in relation to the chloroplasts, for it is in these organs that the starch is localized and undergoes dissolution through the action of the amylase. It is, of course, particularly of the living leaf that knowledge of the amylolytic activity is desired. Important as have been the contributions of Brown and Morris and of other investigators who used similar methods, it is largely because these methods dealt mainly with dried leaves, that it has been necessary to re-examine the subject and to adopt different methods. It must be clearly realized that it is one problem to demonstrate that amylase is present in a certain tissue such as a leaf, but that it is quite another problem to determine quantitatively the activity of this enzyme and to establish the conditions which influence its activity in the living organism. The significance of the difference in point of view will appear in the later discussion and results.

A definite advance in technique was made through the very extensive investigations of Sjöberg (1922) who determined the amylolytic activity of a large number of plants including special determinations of their leaves. Sjöberg used fresh material and thus avoided many of the difficulties inherent in the method of Brown and Morris. His results reveal clearly the tremendous variability in amylolytic activity in different species of plants, of the influence of the age of the plants and of the hydrogen ion concentration of the medium, and serve primarily as a survey of the distribution and activity of amylase in some of the more common plants.

The starch contained in leaves is apparently localized in the chloroplasts. Whether these organisms also contain amylase or whether the amylase is contained solely in the protoplasm or is a part thereof and the relation of the chloroplasts to the protoplasm are problems obviously as yet beyond the reach of chemical methods. The functioning of the chloroplasts is remarkably sensitive to even slight injury, and more drastic treatment, such as with anesthetics, which do not affect enzymes like amylase, has a strong effect on its action in the organized cell. Apparently any influence which affects the colloidal structure of the chloroplasts or the phase relationships of chloroplast and protoplasm greatly inhibits or destroys the normal functioning of the chloroplasts.

The starch in the mesophyll of leaves must be regarded as a transitory reserve food material which is formed and consumed at varying rates depending upon a variety of factors. In order to resolve this rather complex phenomenon into its constituent parts to a degree at least, it has seemed advisable to consider first of all the relatively simple fact of the disappearance of starch from the leaf, that is starch dissolution, independent, for the time being, of the means by which this is accomplished, namely through the activity of the amylase. For convenience of handling the observational facts there are thus two approaches to the problem: quantitative data on the dissolution of starch in leaves and the amylolytic activity of leaves. The first of these is dependent upon relatively simple chemical analysis, the other, a far more complicated matter, is dependent upon the determination of the activity of an enzyme in the leaf tissue and is associated with all the uncertainties inherent in such measurements. The first of these, the starch dissolution, constitutes the description of a phenomenon; the second embodies an attempt to interpret the phenomenon on the basis of the elements which are involved, primarily the activity of an enzyme. Much of the confusion which has existed regarding the phenomenon

has, it would appear, arisen from inadequate analytical methods and modes of experimentation and from a tendency to premature interpretation on the basis of inadequate data.

I. Starch Dissolution in Leaves

That starch dissolution does not take place in leaves in the absence of oxygen was very briefly alluded to by Böhm (1883, p. 51). Similarly, Detmer (1883) and Eisenberg (1907) attempted to show that oxygen is necessary for the formation of amylase in germinating seeds, and Winkler (1898) mentioned that oxygen is necessary for the formation of starch from sugars in leaves. Recently Phillis and Mason (1937) have reported that the formation of starch by leaves, previously destarched and immersed in sugar solutions, is accelerated by oxygen.

Wortmann (1890) also presented some evidence that starch dissolution is affected by the absence of oxygen. Through what must be ascribed to faulty experimental technique, he concluded that leaves contain no amylase, and that starch dissolution must be ascribed to a "direct" action of the protoplasm. In the ripening process of the banana the relatively large amount of starch in the fruit undergoes conversion to simpler sugars. Bailey (1905) found that these changes, incident to the ripening process, proceeded much more slowly when the gaseous exchange was prevented by covering the fruit with paraffin or submerging it in oil. He also obtained some evidence that similar effects are obtained through a deficiency or a complete lack of oxygen. The same general results were reported by Tollenaar (1925), who investigated the conditions affecting the dissolution of starch in tobacco leaves. He concluded that these leaves do not undergo starch dissolution in the absence of oxygen. This he based upon his observation that when the leaves are submerged in water or are confined in an atmosphere of hydrogen or of carbon dioxide no starch dissolution occurred, while in air, at the same temperature, the disappearance of starch could be easily demonstrated. Kern (1928) also noted an inhibiting effect, due to lack of oxygen, in seedlings but not in root tips.

In spite of the fact that the apparent necessity of oxygen for starch dissolution in leaves has stood in the literature of the subject for over half a century; and although it is well known that the activity of amylase, which is presumably responsible for this action, is not dependent upon available oxygen, so far as we have been able to find, no systematic investigation has been carried out on the relation of oxygen to starch dissolution and amylolytic activity of leaves. In any approach to this problem the fact must, of course, be borne in mind that leaves which

have been killed are not capable of hydrolyzing their contained starch, although the killing agents (*e.g.*, freezing, drying, toluene) are of themselves not directly injurious to the amylase. This matter will be discussed in another section of this paper. If, therefore, leaves kept under anaerobic conditions were thereby killed or had suffered serious injury of the chloroplast structure it would not be surprising if no starch dissolution occurred. It would manifestly be essential to keep the leaves under such conditions that they would not be killed or seriously injured by the absence of oxygen.

As a matter of fact some leaves show a remarkable capacity for maintaining their life under anaerobic conditions, though there is an enormous variation in different species in this capacity. Paech (1935) has already called attention to this fact in another connection. The injury to leaves brought about by absence of oxygen is readily detected by the loss of turgor and finally by intercellular infiltration resulting in dark spots, from which the leaf can not recover. It is, however, possible to keep some leaves in an atmosphere of nitrogen or hydrogen for periods of 15 to 24 hours without any signs of injury having occurred, and this is quite a sufficient period for starch dissolution to be determined quantitatively.

Another factor which deserves special consideration is that, so far as can be judged, in previous experiments leaves or other plant parts were *confined* in an atmosphere free of oxygen. There is now evidence that such a procedure may have quite different results from those obtained when leaves are kept in a stream of an irrespirable gas. In a confined and relatively small volume of air starch dissolution of a leaf may be decidedly inhibited even though the concentration of oxygen is considerable or ample for aerobiosis, while in a stream of air starch dissolution proceeds normally. This would indicate that in confined air an accumulation of metabolic products takes place which is deleterious to starch dissolution and that in an air stream these products are constantly removed. It will be shown that the same applies to anaerobic conditions, namely that starch dissolution does occur when some leaves are kept in a stream of nitrogen or hydrogen and that this anaerobic starch dissolution is inhibited in confined atmospheres of these gases, indicating that oxygen is not necessary for dissolution and that this is retarded by the accumulation of respiratory products.

The older observations on the influence of oxygen on starch dissolution have for the most part been made by means of qualitative methods, primarily by the use of the starch iodide test on the leaves. While this has the advantage of simplicity and small amounts of starch

can be detected thereby, the method does not yield quantitative data, so that it can only be said that there is much, little or no starch present. Attempts to make the method more nearly quantitative by the determination of density of the iodine-stained starch granules by means of photography have also not yielded satisfactory results. The problem is, moreover, further complicated by the fact that the various dextrins which are formed as a result of starch hydrolysis give varying degrees of color with iodine ranging from blue, violet, red, orange to colorless. Consequently it is impossible to determine by means of the iodine test how far the hydrolysis of the starch has proceeded.

While, therefore, the effect of various gas mixtures on the starch dissolution of leaves can be demonstrated by means of staining the starch in the leaves with iodine, there is an enormous factor of variability, and little of value as a record can thus be obtained. Consequently we are omitting here the description of the qualitative experiments which led to a quantitative study of this question and shall give only the quantitative results.

Methods

It is perhaps needless to point out that the determination by chemical means of an activity such as starch dissolution in leaves is associated with many experimental difficulties. It can be done only by comparing the amount of starch present before and after a lapse of time, and this naturally involves the destruction of the living tissue. Yet no two leaves, for purposes of comparison, are exactly the same; starch content varies from leaf to leaf and the rate of starch dissolution, and amylolytic activity also varies, especially with leaves of different age and in the leaves of different plants. It is therefore impossible to attain a high degree of accuracy. In order to obtain results which have any true significance at all, reliance must be placed upon sufficiently large samples and selection of the leaf material in such a manner as to reduce to a minimum the effects of individual variability. After considerable preliminary investigation it was concluded that the principle of the half-leaf method offered many advantages which could not be attained in any other way, particularly for leaves of the structure of the sunflower and tobacco. With a sufficiently large number of leaves, five to ten, for each experiment the variation in starch content and in amylolytic activity between the two sets of halves was well within the accuracy of the analytical methods used. In each experiment every effort was made to select leaves as nearly as possible of the same age, from plants of the same age, and grown under as identical conditions

as was feasible. The use of excised leaves has a definite advantage in that the complications arising from the translocation of material between the leaves and the rest of the plant are thus avoided.

The experiments on starch dissolution were carried out with the leaves of *Helianthus annuus* and *Nicotiana Tabacum.* The choice of plant material was determined by several factors: it was essential that the leaves contained sufficient amounts of starch to permit fairly accurate comparative analyses, that the rate of starch dissolution could be conveniently followed experimentally, and that the leaves could withstand anaerobic conditions during the experimental periods. Preliminary experiments showed that all of these conditions were not met by many plants. The leaves of the sunflower represent a medium starch content toward the end of a bright day, 3 to 7 per cent, and a fairly high amylolytic activity; the leaves of tobacco have a high starch content, 5 to 35 per cent, and a low amylolytic activity.

The plants were grown in the greenhouse and in the garden, and leaves from both sources were studied. All of the experiments were done with excised leaves in order to avoid the complications of migration of carbohydrates from and to the rest of the plant. The plants were kept well watered for at least 24 hours previous to cutting of the leaves. This was found to be essential, because the relative amounts of the various carbohydrates contained in the leaves are materially affected by the water content, or the per cent saturation, of the leaves. Comparative analyses at best are difficult to interpret, and water content appears to be one of the most important factors in controlling the ratio of the different carbohydrates present and also has a decided effect on the amylolytic activity as will be shown by special experiments.

The leaves were cut from the plants in the afternoon with the petioles under water, and were allowed to remain in the dark at room temperature for about an hour before starting the experiments. The water used was a mixture of equal parts of boiled tap and distilled water. The leaves were cut along the midrib; the halves without the midrib were immediately prepared for analysis and taken to represent the original condition of the leaf. The other halves, with the midrib, after cutting about a centimeter from the petiole, were immediately placed in the experimental chamber. The chambers used for the starch dissolution experiments in different gas mixtures were the McIntosh and Fildes anaerobic culture apparatus, which were covered to exclude light. The lids of the jars were provided with a metal tube reaching to the bottom for entrance of the gas; the gas was withdrawn from the top of the jar. Wherever rubber tubing was used this was of the sulphur-free variety.

The sealed jars were placed in a water thermostat kept at 30° and immediately the gas, the effect of which was to be studied, was passed through. The hydrogen and nitrogen were first passed over glowing copper and then through water. For the first hour the gasses were passed through the chambers at a rapid rate, about 30 liters per hour, in order to secure uniform composition in the chambers. Efforts to attain this by means of repeated evacuation and flushing of the chambers with the gas proved to be injurious to the leaves, as in the process of evacuation the leaves became infiltrated with water. When nitrogen or hydrogen was used it is, of course, probable that the leaves retained some oxygen, but this amount must certainly have been very small and have been rapidly consumed, for it was not perceptible in the analysis of the gas withdrawn. After the first hour the rate of the gas stream was reduced to about three liters per hour. Samples for the analysis of the gases were drawn from the gas stream after it has passed through the leaf chamber. The experiments were run for 13 to 24 hours and varied according to the species of plant used. Some of these showed injury under anaerobic conditions much more rapidly than others. This injury became evident in loss of turgor, intercellular infiltration, the development of dark spots, etc. It was a matter of surprise, however, how long some leaves, *e.g.*, sunflower, were able to survive a completely anaerobic atmosphere. After termination of the experiment the leaves were removed from the chambers, the midrib cut away, and the leaves weighed.

The leaves were prepared for analysis by first killing through exposing them to the vapors of chloroform. This was done by placing them in a container with a false bottom of wire gauze below which was a mat of filter paper soaked in chloroform. It is important that the leaves should not be allowed to remain too long in the chloroform, that is, beyond the time in which they have been killed and become thoroughly flacid. Depending upon the species this requires about 10 to 15 minutes. If they are permitted to remain too long they exude sap, containing sugars and other soluble material, which if lost naturally causes an analytical error, as was found by special experiments. In leaves with a relatively low water content and thick structure this source of error may be less significant than in thin leaves and those of high water content. The chloroform treatment has the advantages that the leaves in this condition can be dried very much more rapidly than living leaves, and moreover, amylolysis does not take place in leaves killed in this manner. For obvious reasons it is, of course, not desirable to bring the leaves in contact with the liquid chloroform. The loss of starch

in even rapid drying of living leaves is considerable, as is now well established by the work of Molish (1921), Schroeder and Horn (1922) and of Tollenaard (1925), which will be discussed later. This effect may cause serious error in the analysis of leaves for carbohydrates. The drying was carried out in a stream of air at 45–50°. In this manner the leaves could be dried in 1.5 to 3 hours. They were then placed in vacuo over calcium chloride for 24 hours and weighed for the dry weight. Thereafter they were ground in a pebble mill and preserved for the analyses. These were carried out as rapidly as convenient.

For the determination of starch the leaf powders were first dried in vacuo over phosphorus pentoxide for twenty hours and samples of 0.25 to 1 gram were used. The method of starch determination has already been described by Spoehr and Milner (1936). It is, of course, impossible to claim a high order of accuracy for the analysis of material such as leaves; in a large measure the precision and reproducibility depend upon the skill and experience of the analyst. An indication of the order of precision obtainable by this method of starch determination is given in Table I, showing the analysis of the halves of the same leaf.

TABLE I

COMPARATIVE ANALYSES SHOWING THE STARCH CONTENT OF THE HALVES OF THE SAME LEAF

Leaf Material	Per Cent Starch of Dry Material	Per Cent Dry Weight
Tobacco,		
from greenhouse, young	21.32	15.98
	21.92	16.06
from greenhouse, mature	9.72	22.11
	10.14	22.56
Sunflower,		
from garden, mature	4.54	17.71
	4.41	17.82
from greenhouse, mature	5.91	19.15
	5.95	19.27

There is an advantage in drying the chloroformed leaves at 45–50° rather than at 100°. At the latter temperature there occurs a loss of starch, probably through hydrolysis by means of the plant acids, and apparently other materials are also lost at the higher temperature. In Table II are given comparative analyses of halves of leaves dried at 45–50° and at 100° which show a loss of starch at the higher temperature.

That no starch dissolution occurs in the chloroformed leaves, even after long periods of time, will be shown in a later section of this paper.

TABLE II

COMPARATIVE ANALYSES OF CHLOROFORMED LEAVES ONE-HALF OF EACH BEING DRIED AT 45–50°, THE OTHER HALF AT 100°

Leaf Material	Dried at	Per Cent Starch of Dry Material	Per Cent Dry Weight
Sunflower, old	45–50°	4.90	17.24
Sunflower, old	100°	3.46	16.20
Sunflower, young	45–50°	4.71	19.22
Sunflower, young	100°	3.05	18.92
Sunflower, young	45–50°	4.87	25.68
Sunflower, young	100°	3.34	25.30

Starch Dissolution in Irrespirable Gases

It was, first of all, of importance to determine whether oxygen is essential for starch dissolution, that is, whether it is possible for starch dissolution to take place in leaves which are kept in an atmosphere free of oxygen. From the experimental viewpoint it was manifest that different leaves and, in fact, different plant tissue would yield widely varying results, because of differences in their reaction to anaerobic conditions. The complexity of the reactions taking place in a plant tissue under anaerobic conditions and the manner in which they may combine to cause death have been discussed by Grünberg (1932) and need not be gone into here. Some leaves are very much more resistant to these conditions than are others. So also do they show great variation in amylolytic activity or rate of starch dissolution even under aerobic conditions. If a leaf with low amylolytic activity is also rapidly killed under anaerobic conditions it was hardly to be expected that it would be possible to find any starch dissolution when such a leaf was placed in an atmosphere free of oxygen, because starch dissolution does not proceed after death. On the other hand, it was conceivable that in leaves which could remain alive for some time under anaerobic conditions and also possessed a relatively high amylolytic activity, starch dissolution could be detected under these conditions. It was, moreover, important that the starch content of the leaves should be followed quantitatively. This, it must again be emphasized, is impossible by means of the iodine staining method.

The leaves of sunflower and of tobacco are capable of starch dissolution in an atmosphere of nitrogen and of hydrogen as is evident from the

results given in Table III. In these experiments the gases were supplied in a continuous stream during the course of the experiment. Longer periods of time could not be used because the leaves begin to show signs of injury in nitrogen and hydrogen. Carbon dioxide is often toxic even for much shorter periods of time than those given in the table. In the experiments cited in Table III the leaves in carbon dioxide had been killed; there was also no appreciable starch dissolution in these leaves. Under these conditions it was frequently observed that the per cent of starch in the dried leaf material of leaves killed in this manner was higher than in the original condition. This it is believed can be ascribed to the fact that when leaves are killed in this manner, sap containing dissolved matter is lost and the remaining solid material consequently shows a higher percentage of starch.

TABLE III

STARCH DISSOLUTION IN STREAMING HYDROGEN, NITROGEN AND CARBON DIOXIDE, BEFORE AND AFTER BEING KEPT IN THESE GASES AT 30° IN THE DARK, SHOWN AS PER CENT OF STARCH OF THE DRIED LEAVES

Plant	Gas	Time, Hrs	Original		Final		Per Cent Starch Decrease
			Per Cent Starch	Per Cent Dry Weight	Per Cent Starch	Per Cent Dry Weight	
Sunflower..........	H_2	24	3.67	20.09	1.12	19.23	69
Tobacco.	H_2	23	5.01	17.59	3.45	16.41	31
Sunflower...........	N_2	24	4.53	24.06	0.00	19.09	100
Tobacco............	N_2	24	4.68	16.62	2.28	14.92	51
Sunflower....	CO_2	23	2.88	24.47	2.79	24.72	3
Tobacco......... ...	CO_2	23	17.77	22.31	18.27	22.03	0

Although there appears to be no doubt that the leaves of sunflower and of tobacco are capable of starch dissolution under anaerobic conditions in nitrogen and in hydrogen, the action is considerably slower than under aerobic conditions. This fact is brought out by the results of experiments given in Table IV. In these experiments two sets of leaves, as nearly comparable as possible, were used in the manner already described. That is, the halves of each of two sets of six leaves were used for analysis to represent the original condition. The remaining halves of each set were placed in the chambers, one set in air, the other set in the gas the action of which was to be compared. After the time indicated, these halves were also analyzed. The starch dissolution in nitrogen and in hydrogen was roughly one-half that in air.

It should be made perfectly clear that it is unfortunately impossible to compare the rates of starch dissolution as shown in Table III with

those in Table IV, for example. The rate of starch dissolution depends upon many factors and varies with the age of the leaves and of the plant, weather conditions and many other factors. It was, therefore, possible to arrange experiments only in such a manner that certain groups of experiments were quantitatively comparable. Thus each group shown in Table IV is as comparable as is possible, but numerically the values of these can not be compared with those shown in Table III.

TABLE IV

Starch Dissolution in Leaves in Hydrogen, Nitrogen, Oxygen and Carbon Dioxide as Compared with Air, Shown as Per Cent of Starch in the Dried Leaves, Before and After being Kept in these Gases at 30° for the Time Indicated in the Dark

Plant	Gas	Time, Hrs	Original		Final		Per Cent Starch Decrease
			Per Cent Starch	Per Cent Dry Weight	Per Cent Starch	Per Cent Dry Weight	
Sunflower...........	H_2	18	4.10	19.99	2.66	16.67	35
Sunflower...........	Air	18	4.55	20.76	0.81	16.76	82
Tobacco............	N_2	24	8.43	20.93	7.26	19.29	14
Tobacco............	Air	24	8.42	21.13	6.04	16.51	28
Tobacco............	H_2	23	4.46	18.60	3.12	18.29	30
Tobacco............	Air	23	4.24	18.53	0.14	14.57	97
Sunflower...........	O_2	18	4.84	19.21	0.62	15.23	87
Sunflower...........	Air	18	4.74	19.59	0.93	15.38	80
Sunflower..........	CO_2	24	2.68	15.82	2.79	17.82	—
Sunflower...........	Air	24	3.08	15.70	0.15	13.31	95

This does not, however, militate against the conclusions drawn. It is evident that a certain amount of starch dissolution does occur when the leaves are kept in a streaming atmosphere of nitrogen and of hydrogen and that this is less than when the leaves are in air. In an atmosphere of carbon dioxide there is, however, no starch dissolution.

Effect of Streaming and of Confined Gases

So far as can be judged from the published results of earlier investigators the effect of irrespirable gases on starch dissolution of leaves was studied by keeping them in confined atmospheres of these gases for a period of time and then testing the leaves for the presence of starch. Obviously under these conditions, if the leaves remain alive, the composition of the atmosphere surrounding the leaves will be constantly changing, so that toward the end of the experiment the leaves

will not be, for example, in an atmosphere of nitrogen, but in one of this gas plus carbon dioxide. It having already been shown above that no starch dissolution occurs in an atmosphere of carbon dioxide, it was important to determine whether there was a difference in the rate of starch dissolution in a confined atmosphere as compared with one from which the products of respiration were constantly being removed by maintaining a stream of this gas.

TABLE V

Starch Dissolution in Leaves in Streaming (Str.) and in Confined (Con.) Gases, Before and After being Kept under the Conditions Indicated, at 30° in the Dark, Shown as Per Cent of Starch of the Dried Leaves

Plant	Gas	Time, Hrs.	Original		Final		Per Cent Starch Decrease
			Per Cent Starch	Per Cent Dry Weight	Per Cent Starch	Per Cent Dry Weight	
Sunflower..........	Str. air	20	5.00	16.15	0.55	13.11	89
Sunflower...........	Con. air	20	4.66	15.56	1.47	13.31	69
Sunflower...........	Str. air	22	5.28	21.52	1.34	18.17	75
Sunflower...........	Con. air	22	5.29	21.60	2.70	16.93	49
Tobacco...........	Str. air	24	6.04	19.29	0.17	13.48	97
Tobacco...........	Con. air	24	7.00	19.19	2.66	14.34	62
Tobacco............	Str. H_2	18	5.50	19.53	2.87	15.67	48
Tobacco............	Con. H_2	18	5.12	18.78	4.69	18.27	10
Sunflower...........	Str. H_2	23	2.92	20.32	0.00	17.40	100
Sunflower...........	Con. H_2	23	3.07	21.97	1.30	19.65	57
Sunflower...........	Str. O_2	22	4.59	22.92	2.89	19.45	37
Sunflower...........	Con. O_2	22	4.27	22.54	3.03	18.95	29

In Table V are given the results of experiments showing the starch dissolution of leaves in a confined atmosphere as compared with that taking place in streaming gases of the same composition. It is easily demonstrable that the starch dissolution of leaves in a confined atmosphere, be this air or an irrespirable gas, is less than when the leaves are kept in a stream of gas. It would also appear that the effect of the streaming gas is greater in the case of the irrespirable gas than in the case of air or oxygen.

Probably the principle difference in conditions so far as the leaves are concerned between the streaming and the confined gases is that in the former some of the products of respiration are constantly removed from the atmosphere surrounding the leaves, while in the confined gases these products are retained and, in fact, are constantly increasing

in concentration. In order to gain more information on this problem, experiments were so arranged that the gases in the chambers containing the leaves could be circulated by means of a pump. In this manner the composition and amount of gas surrounding the leaves were in no wise affected; the same gas was simply circulated within an entirely closed system containing the leaves, at a rate of about 6 liters per hour. By means of this arrangement provision was then also made that the gases could be washed by introducing into the circulating system appropriate reagents which removed certain constituents of the gas surrounding the leaves. By this latter means the leaves were constantly supplied with the same gas minus the materials which had been removed by the reagent. In this manner it was possible to compare the effect of confined, simply circulating, circulating plus washing, and streaming gases on the starch dissolution of the leaves in the chambers. Some results of this type of experimentation are shown in Table VI,

TABLE VI

COMPARATIVE STARCH DISSOLUTION IN SUNFLOWER LEAVES IN STREAMING (STR.), CONFINED (CON.) AND IN CIRCULATING (CIR.) GASES, BEFORE AND AFTER BEING KEPT UNDER THE CONDITIONS INDICATED, AT 30° IN THE DARK, SHOWN AS PER CENT OF STARCH OF THE DRIED LEAVES

No.	Gas	Time, Hrs.	Original		Final		Per Cent Starch Decrease
			Per Cent Starch	Per Cent Dry Weight	Per Cent Starch	Per Cent Dry Weight	
1	Con. air	22	4.59	19.60	1.62	17.99	65
	Cir. air	22	5.00	20.05	1.83	16.50	63
2	Con. air	22	4.37	21.57	1.85	17.59	58
	Cir. air + KOH	22	4.27	21.78	1.07	17.74	75
3	Con. air	22	3.80	21.14	1.73	17.64	54
	Cir. air + $CaCO^3$	22	3.72	21.16	1.22	17.32	67
4	Str. air	22	3.24	19.71	0.82	16.23	75
	Cir. air + KOH	22	3.51	19.77	0.97	16.48	72
5	Con. H_2	22	3.70	22.18	3.08	20.33	17
	Cir. H_2 + KOH	22	3.26	22.28	2.08	18.03	36

wherein "confined" means that the leaves were kept sealed in the chamber during the entire period of the experiment; "circulating" means that the gas in the chamber containing the leaves was moved by means of a small pump from the exit tube of the chamber, through the pump and again into the chamber. Into the latter system could also be inserted a wash-tube containing, for example, potassium hy-

droxide. "Streaming" means that a continuous stream of fresh gas was passed through the chamber containing the leaves.

It was first of all demonstrated that the mere movement of the gases surrounding the leaves had no influence on the starch dissolution. That is, no appreciable differences in starch dissolution were observed between leaves which were kept in a confined atmosphere as compared with those kept in an atmosphere which was merely circulated by means of the pump. This is shown through the example in Table VI, No. 1.

If, however, a solution of potassium hydroxide was introduced into the circulating system, so that the gas contained in the leaf chamber was continuously circulated through this solution and again passed over the leaves, an appreciable difference in the rate of starch dissolution of the leaves was observable as compared with those which were entirely confined in the same gas. Thus, as can be seen in the example cited in Table VI, No. 2, the starch dissolution of the leaves in confined air was less than that of the air circulated through the chamber and washed with potassium hydroxide. Similarly, the leaves confined in hydrogen showed less starch dissolution than those confined in an atmosphere of hydrogen which was circulated and washed with potassium hydroxide. Moreover, it would appear that washing the circulated air with potassium hydroxide produced conditions which enable the starch dissolution of the leaves to proceed at approximately the same rate as in a stream of fresh air. Thus, the rate of starch dissolution in streaming air is very nearly the same as that of the leaves which were in confined air which was continuously circulated and washed with potassium hydroxide. Table VI, No. 4.

Effect of Carbon Dioxide on Starch Dissolution

It is now generally recognized that the effects of anaerobic conditions on plant cells are of a very complicated nature and the causes leading to the injury or death of the cells are varied, depending upon a number of circumstances (Grünberg 1932). Whether such injury is due to an accumulation of the products of anaerobic respiration, such as carbon dioxide, ethyl alcohol and acetaldehyde or to some form of inanition apparently depends upon particular conditions. In the foregoing experiments the leaves contained an ample supply of carbohydrates. Under anaerobic conditions, that is, in atmospheres of nitrogen or hydrogen, qualitative tests for ethanol and acetaldehyde were obtained in the water of leaf chambers in which the leaves had stood. It seems clear, however, that the removal of carbon dioxide from the atmosphere

tends to accelerate the rate of starch dissolution under aerobic as well as under anaerobic conditions. Consequently the conclusion seems justified that the accumulation of carbon dioxide, as a product of respiration, in the atmosphere surrounding the leaves is detrimental to amylolytic activity of the leaves. In what manner the carbon dioxide exerts this inhibiting influence is, of course, still an open question. That carbon dioxide when added to gas mixtures has an inhibiting influence on the rate of starch dissolution of leaves is demonstrated by the following series of experiments.

In these experiments two sets of leaves, as closely similar in age and amylolytic activity as possible, were used; in one set the rate of starch dissolution was determined in air, in the other in air to which carbon dioxide had been added. One-half of each leaf from each set was used for the determination of the starch, as representing the original condition, the other halves were placed in the chambers for 22 hours, at 30°, with, in one case a stream of air, and in the other case a stream of air plus carbon dioxide passing through the leaf chambers. At the end of the experiment these leaves were analyzed in the manner described. In Table VII are given the results of these analyses in terms of the ratio

TABLE VII

THE EFFECT OF DIFFERENT CONCENTRATIONS OF CARBON DIOXIDE ON THE STARCH DISSOLUTION IN SUNFLOWER LEAVES

The first line gives the concentrations of carbon dioxide in air; the second line the ratios of the per cent decrease of starch in leaves, during 22 hours, kept in air plus carbon dioxide to those kept in air. All gases were streaming and the temperature was 30°.

Per Cent CO_2	97	88	49	25	21	17	12	10	7	5	3
Ratio of per cent starch decrease in CO_2 + air to decrease in air. .	.45	.48	.46	.77	.85	.73	.88	.51	.60	.91	1.0

of the decrease of starch in the two sets of leaves for each concentration of carbon dioxide. For reasons already explained, it is impossible to compare the rates of starch dissolution per se in a long series of experiments of this nature. It is felt that the ratios of the per cent starch decrease of the leaves in air to those in air plus carbon dioxide give a fairer comparison.

It is manifest that carbon dioxide had a definitely inhibiting influence on the rate of starch dissolution of these leaves. It should be pointed out that this effect is undoubtedly a complicated one. In this series of experiments concentrations of carbon dioxide above 20 per cent were clearly toxic, so that leaves in these concentrations of carbon dioxide

showed signs of injury. In the higher concentrations of carbon dioxide the leaves were probably killed quite rapidly. Below concentrations of carbon dioxide of 20 per cent the sunflower leaves evidenced no sign of injury, although other species, such as tobacco and sugar beet, could not withstand even these concentrations without injury for as long as 20 hours, such periods being necessary to secure quantitative measurements of starch dissolution. It should be mentioned at this point that in other series of experiments with sunflower leaves, of a later planting, no signs of injury were observable in concentrations of carbon dioxide of 45, 60 and even 80 per cent.

A similar effect of carbon dioxide on the starch dissolution is also noticeable when the leaves are kept under anaerobic conditions. Thus leaves kept in a stream of hydrogen plus carbon dioxide showed a lower rate of starch dissolution than similar leaves which had been kept in a stream of hydrogen alone, as is shown in the results in Table VIII. It is

TABLE VIII

STARCH DISSOLUTION IN SUNFLOWER LEAVES IN STREAMING HYDROGEN AS COMPARED WITH STREAMING HYDROGEN PLUS CARBON DIOXIDE, BEFORE AND AFTER BEING KEPT UNDER THE CONDITIONS INDICATED, AT 30°, IN THE DARK, SHOWN AS PER CENT OF STARCH OF THE DRIED LEAVES

Gases	Time, Hrs.	Original		Final		Per Cent Starch Decrease
		Per Cent Starch	Per Cent Dry Weight	Per Cent Starch	Per Cent Dry Weight	
H_2	23	3.97	22.85	1.18	19.01	70
$H_2 + 3.6\%\ CO_2$	23	4.00	21.56	2.13	19.26	47
H_2	23	4.57	22.66	0.32	18.47	93
$H_2 + 10\%\ CO_2$	23	5.00	23.66	2.11	20.35	58

possible that in previous experiments the conclusion that leaves do not undergo starch dissolution under anaerobic conditions may, in part at least, be ascribed to this effect, inasmuch as in the older experiments the leaves were confined with irrespirable gases under conditions which permitted the accumulation of carbon dioxide and other products of respiration.

II. AMYLOLYTIC ACTIVITY OF LEAVES

The foregoing experiments on starch dissolution in leaves were based upon the determination of the disappearance of starch under various conditions. This presumably occurs through the action of amylase. But the interpretation of the results of a study of starch dissolution from the point of view of amylolytic activity is a far more

difficult task. This is largely due to the fact that there are still many uncertainties in attempting to correlate starch dissolution with such determinations as we are capable of making of the amylolytic activity of leaves. It must be admitted that the methods of determining the amylolytic activity of a plant tissue probably give only an approximation of this activity as it actually exists in the living cell. There is, for example, the rather important difference that amylolytic activity is usually determined by means of the hydrolytic activity of the leaf preparation on gelatinized starch, while in the living cell the dissolution of the organized starch granule is involved. The leaf preparations or extracts act very slowly upon uninjured starch granules, if at all. Apparently the initial stages in the dissolution of starch granules require the living protoplasm, or some component thereof, which is destroyed when the cell is killed. So that it is possible that amylase preparations and killed leaf material contain only fragments of the entire enzyme complex which is in the living cell and is required to break down the starch granules as they exist in the living cell. There is furthermore the fact that the amylolytic activity of the leaf cells can increase with surprising rapidity. Whether this is due to the elaboration of more amylase by the protoplast or to the rapid change of some condition within the protoplasm favorable to the action of the enzyme, it is impossible to say.

Nevertheless, there can be but little doubt that the weight of evidence is definitely in favor of the conclusion that the dissolution of starch in the chloroplasts is conditioned by a starch dissolving enzyme. The reliability of conclusions concerning this amylolytic activity in leaves depends in a large measure upon the precision of the analytical methods and the exact maintenance of the essential conditions for obtaining reproducible analytical results. During the preliminary stages of this investigation the determinations of amylolytic activity of duplicate samples of leaf material showed variations as great as 40 per cent by following methods described in the literature. It was necessary to establish conditions and to adhere to these rigorously in order to obtain even fairly satisfactory results. Finally a procedure was developed by means of which a reproducibility of $\pm$ 2 per cent was obtained.

Method of Determining Amylolytic Activity

A weighed sample of fresh leaf material was kept 30 min. with 1 cc. of toluene in a glass-stoppered container. The killed material was ground for 10 min. with half its weight of quartz sand in a glass mortar. Water or 4 per cent phosphate solution was added to obtain the desired fluidity for grinding. The reaction mixture contained the ground

leaf material plus 5 cc. 4 per cent phosphate solution per 100 cc. plus a solution of soluble starch to make 1.00 gm. starch per 100 cc. plus water to make to volume. The starch solution was made from Merck's "soluble starch" and was boiled for 3 min. to insure complete gelatinization. Most determinations of amylolytic activity were made upon 2–5 gm. samples of leaf material in 250 cc. reaction mixture, using 4 per cent phosphate solution of pH 6.00. The reaction mixture was transferred to a glass-stoppered flask, 1 per cent of its volume of toluene added, and placed in a water thermostat kept at 30°. At intervals portions of the reaction mixture were withdrawn, centrifuged, and 5 cc. samples of the supernatant liquid used for the determination of the reducing power by means of the method described by Spoehr (1919).

Sjöberg (1922) reports amylolytic activities as *Sf* (Verzuckerungsfähigkeit):

$$Sf = \frac{k \cdot \text{gm. starch}}{\text{gm. preparation}},$$

where k is the monomolecular reaction constant,

$$k = \frac{1}{t} \log \frac{a}{a - x},$$

where t is the time in hours, a is the quantity of starch initially present and x is the quantity of starch hydrolyzed.

As has been pointed out by Euler and Svanberg (1921) and by Euler and Josephson (1923) the activity of an enzyme preparation can be expressed by the generalized formula:

$$Xf = \frac{k \cdot \text{gm. substrate}}{\text{gm. enzyme preparation}},$$

when k is proportional to the original amount of substrate. Here k is the monomolecular velocity constant for each determination and holds only in the early stages of the reaction. The limits within which this equation holds are also defined by Euler and Svanberg (1921) as are the conditions under which this equation can be applied.

Inasmuch as we were primarily interested in obtaining comparative values for the rates of hydrolysis of starch by different species of leaves, the above generalized formula was applied in the form used by Sjöberg. In order, however, for Sjöberg's formula to be applied rigorously, the maximum amount of starch hydrolyzed by each enzyme preparation would have to be established for each determination. For the present purposes this was not deemed to be essential.

All of the amylolytic activities in this report are calculated from the reducing powers after a reaction time of 24 hours at 30°. In calculating our k, $t = 24$, $x =$ cc. of Benedict's solution used in the 24 hour reducing power determination, and $a =$ calculated cc. of Benedict's solution if all of the starch in the reaction mixture had been hydrolyzed to maltose, although this probably does not meet the exact conditions stipulated by Euler and Svanberg. In this report the symbol A (amylolytic activity) is used instead of Sjöberg's *Sf*. The reproducibility of the method was determined by twelve experiments in each of which duplicate samples of different leaf materials were used.

It is, of course, essential to make preliminary determinations of the activity of unknown material in order to obtain an approximate value, because the amount of sample necessary for a determination and the amounts of reagents required must of necessity differ according to the activity, if comparable and reliable results are to be obtained.

The results which may be obtained by the determination of amylolytic activity depend in so large a measure upon the manner in which the plant material is treated that it seems advisable to present below in very brief form some of the experiments upon which are based the method which was finally adopted.

There appears to be little doubt that many of the conflicting results which have previously been obtained on the amylolytic activity of plant tissue are due to insufficient consideration of the importance of a standard method of preparing the plant material. Very little comparative

TABLE IX

Comparative Amylolytic Activity of Sunflower Leaf Material Treated in Three Different Ways for Analysis: *A*, Ground at Once; *B*, Stood for 30 Minutes with Toluene Before Grinding; *C*, Dried Immediately in Vacuum over $CaCl_2$ and Ground with the Addition of Water

Treatment	*A*	Relative Activity
A	0.009711	88
B	0.011070	100
C	0.008169	74
A	0.004784	69
B	0.006975	100
C	0.002219	32

study has been given to this aspect, although the results may differ considerably depending upon the manner in which the leaves have been treated. This may be illustrated by the examples given in Table IX, where are shown the comparative amylolytic activity of triplicate

samples of the same sunflower leaf material, treated in three different ways, and showing the great variance in the final results.

Palladin and Popoff (1922) have shown that when leaves are permitted to undergo autolysis in the presence of toluene or chloroform there results apparently a liberation of amylase. The arrangement of Palladin and Popoff's experiments were, however, not such that very much can be learned from them of a more precise nature. It is already clear from the experiments such as are cited in Table IX that when the leaves are allowed to stand with toluene the amylolytic activity is greater than when they are ground immediately. From our results it would appear doubtful that this is due to an autolytic process which frees amylase from the protoplasm, as was assumed by Palladin and Popoff. For when leaves were allowed to remain for longer periods of time with toluene, that is, of about the same order of duration as in the experiment of Palladin and Popoff, there occurs an unmistakable decrease in the amylolytic activity. In Table X are shown the results

TABLE X

Comparative Amylolytic Activity of Sunflower Leaf Material With and Without Toluene, and Treated with Toluene for Different Lengths of Time

Treatment	A	Relative Activity
Ground fresh	0.004330	63
30 min. with toluene	0.006844	100
180 min. with toluene	0.006844	100
60 min. without toluene	0.003666	54
30 min. with toluene	0.004556	100
1 day with toluene	0.003666	81
3 days with toluene	0.002892	64
6 days with toluene	0.002318	51
12 days with toluene	0.001861	41
24 days with toluene	0.001634	36

of two series of experiments which demonstrate the effect of the toluene treatment as compared with untreated leaves, and also the effect of prolonged treatment with toluene on the amylolytic activity. In each series of these experiments identical samples of sunflower leaves were used. The treatment with toluene was carried out so that each sample was treated with one cc. of toluene in a glass-stoppered flask and kept in the dark at room temperature for the length of time indicated.

It is, of course, quite impracticable to attempt to obtain accurate results of the amylolytic activity of leaves from dried samples. This is due to the fact that during drying, without previous killing, the leaf undergoes drastic changes with respect to its amylolytic activity.

Brown and Morris (1893) were of the opinion that "It is only by *drying* the leaf that the full activity of the contained diastase can be appreciated. If a leaf is air-dried at about 40–50°, then finely powdered, and put under favorable conditions for hydrolysing starch, the hydrolytic action is very much greater than that of the freshly pounded undried leaves." Such is undoubtedly the case. But this procedure leaves out of account the fact that in the process of losing water through transpiration the leaf gains substantially in amylolytic activity. This phenomenon will be considered in more detail in a later portion of this paper. Because of this gain in amylolytic activity during water loss it would appear manifest that material prepared in this manner can give no reliable information concerning the activity of the leaf in the fresh condition. And from the biological point of view that is the important consideration.

Moreover, leaves which have been first killed by means of toluene and are then dried also show a loss of amylolytic activity, although this is less than the loss sustained when the leaves are stored with toluene in the fresh condition. Below are given some results of determinations of amylolytic activity of identical samples of sunflower leaves which were treated with toluene for 30 minutes and then dried in vacuo over calcium chloride. The relative amylolytic activities, determined in the identical manner as in the experiments cited above, were after drying 1 day, 100; after 2 days, 99; after 4 days, 91; and after 8 days, 87.

The decrease in amylolytic activity of stored leaf material is not due to oxidative changes. Special experiments to determine this point have shown that the decrease occurs in leaf samples which were stored in hydrogen quite as rapidly, and in some cases even more rapidly, than those stored in air. This applies to leaf material killed with toluene and stored in hydrogen in the moist condition and also to material, killed in an atmosphere of hydrogen in the same manner and dried and stored in hydrogen.

Toluene has certain advantages over chloroform for killing leaf material. In the use of chloroform there is always the danger of traces of hydrochloric acid, which may affect the hydrolysis of starch directly and also be of influence on the hydrogen ion concentration of the reacting mixture and thus affect the activity of the amylase.

Amylolytic Activity of Different Species

Sjöberg (1922) has already called attention to the great variation in amylolytic activity of mature leaves of different species of trees. It will be shown that the same applies to different species of herbaceous plants.

However, in view of the fact that the same species may show a variation of many fold depending upon the age of the plant, the conditions under which it was grown, etc., single determinations of the amylolytic activity of different species of plants mean very little, as will be evident from the results shown in Table XI. In the cases cited where the

TABLE XI

THE AMYLOLYTIC ACTIVITY OF THE LEAVES OF DIFFERENT SPECIES OF PLANTS

The relative values of the mean activities are summarized by assigning the arbitrary value of 1 to the least active species and expressing the others in proportion.

Species	No. of Determinations	Amylolytic Activity (A)			Comparative (A) Mean
		Maximum	Minimum	Mean	
Trifolium repens	7	.02410	.006480	.01267	358
Beta vulgaris	7	.01414	.003245	.008982	254
Helianthus annuus	143	.01429	.001105	.007940	224
Achillea millefolium	6	.009441	.003553	.005718	161
Tropæolum majus	5	.003497	.002060	.002849	80.5
Iris germanica	4	.002823	.0009112	.001843	52.0
Chlorella pyrenoidosa	1	—	—	.001416	40.0
Nicotiana Tabacum	11	.002019	.0008080	.001308	36.9
Hordeum sativum	2	.001195	.001103	.001149	32.4
Saxifraga crassifolia	3	.0004226	.0004109	.0004162	11.8
Cladophora sp.	1	—	—	.0002063	5.83
Pelargonium zonale	1	—	—	.0001833	5.18
Begonia luminosa	3	.0001894	.00008349	.0001250	3.53
Elodea canadensis	1	—	—	.0001138	3.21
Mesembryanthemum spectabile	1	—	—	.00005796	1.64
Sedum præaltum	1	—	—	.00005211	1.47
Rumex scutatus	4	.00005833	.000008583	.00003541	1.00

number of determinations and the activities are small the values given for the mean activity are but an indication of the degree of difference in the amylolytic activity of the several species. It is striking, however, that the acid plants examined, such as begonia, sorrel, sedum and mesembryanthemum, show a remarkably low amylolytic activity. It is as yet impossible to account for the very wide variations obtained in the values for the activities of the same species and little more can be said beyond that age and the previous treatment of the plant as to water-content and transpiration, temperature and probably also illumination have a definite effect on the amylolytic activity of the leaves.

If there is a relation between the amylolytic activity of a leaf and its starch forming power, or at least its starch content, as was suggested by Brown and Morris (1893, p. 643), it is of a much more complicated nature than they apparently supposed. This lack of correlation between amylase activity and starch content in leaves has already been

pointed out by Eisenberg (1907) on the basis of starch tests made with the iodine method and a very uncertain method of determining amylolytic activity.

The variability in amylolytic activity of the same species of leaves, shown in Table XI, is due to a number of factors. It has been known for some time that the activity varies considerably with the age of the leaves, and this has been amply confirmed. This may be illustrated by the following example of sunflower leaves of different age taken from the same plant. Leaf "A" was the oldest one which had not yet turned yellow; leaf "D" was the youngest one which was large enough to serve as an analytical sample; leaves "B" and "C" were taken from intermediate positions. Their relative amylolytic activity was as follows: "A," 76; "B," 86; "C," 100; and "D," 77.

The inconstancy of the amylolytic activity may in a measure also be associated with differences in the time of day at which the leaves are collected. However, there appears to be no consensus of opinion that this is connected with illumination. Brown and Morris (1893) on the basis of what must now be considered as insufficient data concluded that in the leaves of *Tropæolum majus* and of *Hydrocharis Morus-ranae* there is in the dark an increase of amylase with the depletion of starch. Eisenberg (1907) could not confirm this with *Pisum sativum;* he found rather that leaves with high starch content and which were well illuminated showed a higher activity than shade leaves of the same species with little starch. Sjöberg (1922) came to the conclusion that there was no relationship between illumination and amylolytic activity; the irregularities in activity which he found could not be directly correlated with light. In all probability the differences in activity of leaves at different times of the day are due to water relations and to temperature. For this reason it is of considerable importance to pay special attention to these factors and to the age of the leaves in the choice of material for comparative analytical purposes.

Amylolytic Activity of Half Leaves

The determination of the causes for the variation of amylolytic activity in leaves, or perhaps more accurately, how this activity varies with environmental conditions, is a matter of considerable experimental difficulty. It has already been indicated that the variation in activity between individual leaves is so great that the comparison of these at different times or after being exposed to different conditions involves large errors. After a great deal of preliminary experimentation we have concluded that the highest degree of precision can be obtained by

means of the "half-leaf method," that is, by making determinations of the amylolytic activity of fractional parts of the same leaves before and after subjecting these to certain experimental conditions. But even with this method it is quite essential to work out the degree of precision of the method as applied to any particular plant material, and to adhere closely to the details of the method of analysis. The halves of single leaves may exhibit great differences in activity, especially in the case of large leaves, so that it is inadvisable to use too small samples, *i.e.*, too few leaves.

Reproducible analytical results could be obtained by using at least three leaves, in the case of sunflower, rapidly cutting the longitudinal halves of these into squares of about one centimeter, thoroughly mixing these, and using 2.5 grams as a sample for the determinations as described above. Ten sets of three leaves each of sunflower leaves yielded results showing a variation in the amylolytic activity of $\pm$ 3 per cent between the two halves. After cutting the leaves from the plants they were kept in the dark, at 20–22°, with the petioles in water, for three hours, before preparing them for analysis as just described.

The hydrogen ion concentration of the leaf saps was determined on the remainder of the leaf-halves which had been cut into squares. These were frozen at $-$ 80° C. and after thawing the sap was pressed out in a stainless steel container with a pressure of 4,500 pounds per square inch. The hydrogen ion concentrations of the clear, expressed saps were determined by means of a glass electrode and are reported on the pH scale.

Effect of Carbon Dioxide on Amylolytic Activity

In the first section of this paper it has been shown that carbon dioxide in the atmosphere surrounding the leaves exerts an inhibiting effect on the starch dissolution. A similar effect of carbon dioxide is noticeable on the amylolytic activity of leaves. The experiments which demonstrated this were arranged in very much the same manner as those on starch dissolution. For the determination of the effect of each concentration of carbon dioxide two sets of four sunflower leaves were used, one for the determinination of the amylolytic activity of leaves kept in air containing a certain percentage of carbon dioxide, the other set for a comparative determination of the amylolytic activity in air alone. The leaves were selected to be as nearly identical in activity as possible. They were kept for three hours in the dark at room temperatue before starting the experiments. One half of each set was used for the determination of the original amylolytic activity, the

remaining halves of each set were placed in the leaf chambers and a stream of air plus carbon dioxide passed through one and of air through the other, the chambers being maintained at 30° for 23 hours. At the end of this time the amylolytic activity was determined in each set as described.

The effect of various concentrations of carbon dioxide on the amylolytic activity is shown in Fig. I. As has been pointed out before,

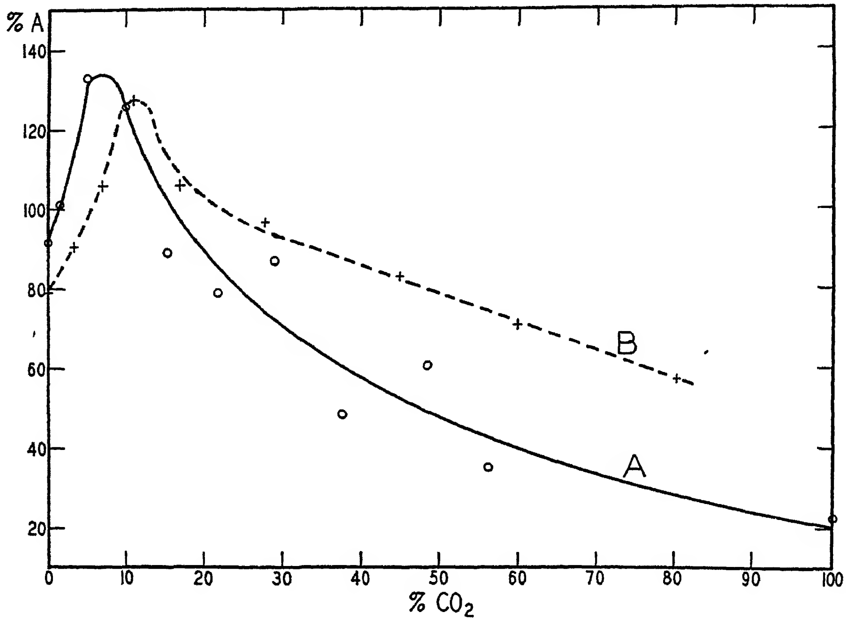


FIG. I. Effect of carbon dioxide upon the amylolytic activity of sunflower leaves. Curve A represents the first, and curve B the second series of determinations.

in a series of determinations of this nature the activity of all the sets of leaves cannot be compared because of variability. Consequently, the analyses have been calculated in terms of per cent activity of the original activity. That is, the activities of the half leaves which were in carbon dioxide plus air were calculated as per cent of the activities of the halves taken at the beginning of the experiment. And similarly, the activities of the half leaves which were kept in air were calculated as per cent of the activities of the halves taken from these at the beginning of the experiment. The ratios of these two percentage values are taken as an indication of the effect of various concentrations of carbon dioxide on the amylolytic activity of the leaves. The ordinate of the curve in Fig. I therefore represents per cent A in CO_2 of original over per cent A in air of original.

Two series of experiments were carried out for this purpose; in the first series leaves were taken from plants growing in the garden June 29 to July 20, in the second series from plants of about the same age during October 7 to 27. In the first series the leaves were killed and wilted in 100 per cent carbon dioxide; in 56, 48, and 38 per cent carbon dioxide the leaves were killed or injured, but not wilted, showing progressively less infiltration with lower concentrations of carbon dioxide. Leaves in 29 per cent of carbon dioxide and less showed no signs of injury. In the second series the leaves showed no signs of injury, astonishingly enough, even in 80 per cent carbon dioxide.

Both series of measurements show an increase in amylolytic activity at lower concentrations of carbon dioxide and a decrease at higher concentrations. In the first series the inhibiting effect becomes apparent at about 25 per cent, while in the second series at about 50 per cent.

It has already been shown by Thronton (1933) and by Fife and Frampton (1935) that the treatment of plant tissues with high concentrations of carbon dioxide results in a decrease of the hydrogen ion concentration of the sap. A similar effect was observed in the sunflower leaves the amylolytic activity of which was here determined after the leaves had been in various concentrations of carbon dioxide. The effect of the carbon dioxide on the hydrogen ion concentration of the sap of the leaves is particularly pronounced with concentrations above about 15 per cent by volume of carbon dioxide. In these experiments the leaves were exposed to the carbon dioxide for 23 hours; at the end

TABLE XII

THE EFFECT OF VARIOUS CONCENTRATIONS OF CARBON DIOXIDE ON THE HYDROGEN ION CONCENTRATION OF THE SAP OF SUNFLOWER LEAVES

Per Cent CO_2 in Air	pH of Sap of Leaves Kept in CO_2 Plus Air		pH of Sap of Leaves Kept in Air	
	Original	Final	Original	Final
	1	2	3	4
100	6.98	7.90	7.00	7.05
56.2.......	6.88	7.28	6.89	6.88
48.6.......	6.80	7.53	6.93	6.90
37.6.......	6.95	7.12	6.90	6.85
29.0.......	6.80	7.06	6.80	6.73
21.8.......	6.75	6.96	6.78	6.72
15.3.......	6.88	6.88	6.83	6.75
10.0.......	6.82	7.07	6.83	6.88
5.0.......	6.87	6.87	6.92	6.92
1.4.......	6.75	6.73	6.75	6.71
0.......			6.90	6.92

of this time the leaves in the control experiments in air showed only very slight differences in hydrogen ion concentration as compared with that at the beginning of the experiment. In Table XII are shown the values of the pH determinations of the four sets of leaf-halves, namely *1*, showing the original condition of the leaf-halves to be kept in various concentrations of carbon dioxide; *2*, of the leaf-halves which had been kept in these gas mixtures for 23 hours; *3*, showing the original condition of the leaf-halves to be kept in air for 23 hours; and *4*, of the leaf-halves which had been kept in air for this period.

The decrease in amylolytic activity in general follows the decrease in hydrogen ion concentration of the expressed sap. Although it is recognized that the measurements of the hydrogen ion concentration of the expressed saps may give only an uncertain value of the true condition as it pertains to the protoplasm, it would appear that a change in hydrogen ion concentration due to carbon dioxide is at least one factor responsible for the decrease in amylolytic activity in higher concentrations of carbon dioxide.

The Effect of Hydrogen Ion Concentration on Amylolytic Activity

It is a very familiar fact that the activity of amylase is very decidedly affected by the hydrogen ion concentration of the medium. The amylases from different sources apparently show an optimum at different pH values, which indicates essential differences of the individual enzymes, Oppenheimer (1925). Very little attention has been paid to the leaf amylases, Sjöberg (1922) determined the pH of maximum activity for five species of leaves and found this to be at pH = 5.0–5.4; that is, about the same hydrogen ion concentration as is maximal for malt amylase.

The hydrogen ion concentration of maximum amylolytic activity was determined for a number of different species of leaves. These were selected with the purpose of obtaining as wide a range as possible in the hydrogen ion concentration of their saps. Thus sunflower leaves have a sap of pH 6.7–7.1 and the leaves of Begonia a sap of pH 2. These values were obtained with the expressed sap, and there is manifestly no means of determining whether they in any sense represent the hydrogen ion concentration of the protoplasm, of the chloroplasts or wherever the amylolysis occurs. Nevertheless, the amylolytic activity of these different leaves exhibits a maximum at different hydrogen ion concentrations, indicating that, in a measure at least, they are adjusted to different conditions which parallel the hydrogen ion concentrations of the expressed saps.

TABLE XIII

AMYLOLYTIC ACTIVITY OF GROUND LEAF MATERIAL AT DIFFERENT HYDROGEN ION CONCENTRATIONS

Reaction Mixture pH		Relative A	Reaction Mixture pH		Relative A
0-Hr.	24-Hr		0-Hr.	24-Hr.	
Helianthus annuus			*Trifolium repens*		
5.75	5.65	72.0	3.75	3.75	17.2
6.02	5.87	90.2	4.58	4.60	75.6
6.23	6.16	100.0	4.98	5.00	92.1
6.55	6.45	100.0	5.47	5.45	96.2
6.76	6.68	98.2	5.90	5.88	97.6
6.95	6.90	91.9	6.37	6.38	100.0
7.40	7.38	61.4	6.84	6.85	94.8
7.82	7.83	26.7	7.26	7.25	88.0
Nicotiana Tabacum			*Tropæolum majus*		
5.19	5.50	53.7	5.50	5.53	70.7
5.38	5.56	64.2	5.68	5.68	79.1
5.63	5.73	82.6	5.97	5.90	96.1
5.95	5.97	92.1	6.37	6.23	100.0
6.37	6.33	97.2	6.80	6.70	94.2
6.82	6.76	100.0	7.20	7.00	98.7
7.23	7.18	87.7	7.50	7.36	57.1
7.60	7.54	64.2	7.70	7.60	35.3
Beta vulgaris			*Tropæolum majus*		
5.28	5.48	62.2	5.70	5.72	80.9
5.45	5.59	73.9	6.00	5.96	91.4
5.67	5.75	85.7	6.40	6.30	100.0
6.00	6.02	96.7	6.65	6.53	94.8
6.42	6.42	97.7	6.85	6.71	87.0
6.87	6.85	100.0	7.05	6.91	80.4
7.30	7.26	92.9	7.26	7.02	94.8
7.57	7.67	88.4	7.57	7.37	65.0
Saxifraga crassifolia			*Begonia luminosa*		
4.53	5.13	93	3.22	4.00	66
5.10	5.52	96	3.72	5.28	91
5.37	5.67	100	4.04	5.67	100
5.67	5.73	96	5.33	6.22	87
6.10	6.12	95	6.05	6.48	78
6.53	6.55	94	6.40	6.64	69
6.86	6.87	90	6.62	6.81	60
7.15	7.15	85	6.83	7.00	51
Rumex scutatus					
4.13	5.15	100			
5.12	5.97	100			
6.10	6.46	100			
6.55	6.81	80			
6.86	7.05	60			
6.97	7.13	60			
6.98	7.20	60			
7.00	7.24	53			

In these experiments the leaf material was ground with sand and water only. The ground mass was made to volume (100 or 200 cc.) with water, and aliquot parts of this suspension were put into each of a series of reaction mixtures. The pH of the suspension was measured, but it does not, of course, represent the pH of the leaf sap. The aliquot reaction mixtures were made to different hydrogen ion concentrations. This was done by adding 5 per cent by volume of 4 per cent phosphate solutions of different pH values. To these was added soluble starch solution to make 1.00 gm. starch per 100 cc. reaction mixture. The pH of each reaction mixture was measured at the time of preparation and again after 24 hours at 30°.

The pH of the reaction mixture after 24 hours may or may not be the same as the pH at the time of preparation. Differences as great as 1.0 to 1.6 pH unit were observed in the case of the acid plants, *Saxifraga*, *Begonia* and *Rumex*, whereas with the other species, the differences were 0.2 pH unit or less.

It was noted that it was not possible to predict the pH of the reaction mixture from the pH of the 4 per cent phosphate solution used in its preparation. In two cases, *Begonia* and *Rumex*, the reaction mixtures were more acid than the corresponding phosphate solutions; in the other cases the reaction mixtures were more alkaline. Also, duplicate series of reaction mixtures made with different samples of the same species of leaf may check closely or they may differ by nearly one pH unit.

In Table XIII are given the results of the determination of amylolytic activity at different hydrogen ion concentrations. The pH values shown are the measured values of the reaction mixtures. The amylolytic activity of *Rumex* is so small that the data are of qualitative rather than quantitative nature. In the case of *Tropœolum* two peaks

TABLE XIV

HYDROGEN ION CONCENTRATIONS AT WHICH THE LEAVES SHOWED MAXIMUM AMYLOLYTIC ACTIVITY

Species	pH of Suspension	pH Maximum A
Beta vulgaris L.	7.15	6.86
Nicotiana Tabacum L.	6.68	6.79
Helianthus annuus	9.15	6.50
Trifolium repens	6.98	6.37
Tropæolum majus	6.58	6.35
Rumex scutatus	3.95	5.55 (?)
Saxifraga crassifolia	5.45	5.52
Begonia luminosa	3.20	4.85

were found in the pH-activity curve, the maximum at pH 6.30–6.35 and the second peak at pH 7.10–7.14.

In Table XIV are listed the hydrogen ion concentrations, in terms of pH, at which the various species of leaves examined showed maximum amylolytic activity.

III. Starch Dissolution and Amylolytic Activity in Wilting Leaves

A very interesting aspect of the phenomenon of starch dissolution in leaves is that this is apparently definitely accelerated by decrease of the water content of the leaves. In the investigations of this aspect of the problem by Rywosch (1908), by Molisch (1922), and by Kisselew (1927) conclusions were based entirely upon qualitative tests for starch with iodine, and even in the more quantitative investigations of Schroeder and Horn (1922) and of Tollenaard (1925) no direct quantitative determinations of starch were made. The remarkable fact that the accelerated starch dissolution under conditions of rapid water loss from the leaves results in an increase of sucrose, attaches particular interest to the phenomenon. That the rate of starch dissolution does, in fact, take place more rapidly in leaves which have lost water to the point of wilting as compared with those amply supplied with water, was demonstrated by experiments in which the starch was determined directly, Table XV. At lower temperatures (5°) the difference in the rate of starch dissolution between leaves with and without water is less striking,

TABLE XV

The Effect of Decrease in Water Content on the Starch Dissolution of Leaves, Before and After being Kept in the Dark for 18 Hours, in Air, Shown as Per Cent of Starch of Dried Leaves

Plant Material and Conditions	Original		Final		Per Cent Starch Decrease
	Per Cent Starch	Per Cent Dry Weight	Per Cent Starch	Per Cent Dry Weight	
Sunflower, in water, 22°	4.20	22.11	3.49	19.28	17
Sunflower, without water, 22°	4.33	19.28	1.34	28.47	69
Sunflower, in water, 5°	3.49	20.29	0.82	19.95	77
Sunflower, without water, 5°	3.46	20.03	0.65	26.30	81
Sunflower, in water, 5°	2.47	21.83	0.60	21.93	76
Sunflower, without water, 5°	2.72	22.43	0.55	29.82	80
Tobacco, in water, 22°	3.51	18.47	1.06	14.83	70
Tobacco, without water, 22°	2.45	18.06	0.04	23.27	98

because at lower temperatures there is also an increased amylolytic activity, as will be shown in the following section.

The increased rate of starch loss in the dark in wilting leaves has been ascribed to transpiration, that is, to the water loss from the leaves, rather than to a lower water content. Thus Neger (1919) showed that the leaves of the potato plant undergo more rapid starch dissolution under conditions of rapid transpiration than when this is low, and he ascribes this to an increase in aeration accompanying the high rate of transpiration. Rywosch (1908) found that the cells which were nearest the conducting bundles were the last to lose their starch under conditions in which the transpiration of the leaf exceeded the supply of water. The results of Tollenaard (1925) would indicate that the phenomenon is not due to transpiration but rather to a decrease in the water content of the leaves. Iljin (1923) reported an increase of 40–70 per cent in the rate of respiration of mesophytic leaves during wilting. Iljin (1930) also showed that this increased rate of respiration during wilting has a permanent injurious effect on the leaves and is made evident by a decreased rate of synthesis of organic material. That the increased starch dissolution cannot, however, in all cases be ascribed to an increased rate of carbohydrate consumption through respiration under conditions of water deficit would be indicated from the results of Collorio (1928). From these it would appear that the relations of water content to rates of respiration are very complex. But there is little doubt that water content, or the degree of saturation of the protoplasmic colloids, is of the greatest importance both for the hydrolytic enzymes as well as for the synthetic ones. Investigations of this aspect of the problem, both on the dissolution and on the synthesis of starch, are still in progress. This much seems to be certain, that the formation and dissolution of starch in the plastids is not a simple matter of a starch-sugar equilibrium, as has been commonly assumed. The concept that the formation of starch depends upon a sufficiently high concentration of mono- and di-saccharides, and that when this concentration falls below a certain point, starch dissolution sets in, no longer seems to be tenable. Both the formation and dissolution of starch, and for that matter also the formation of sucrose, are dependent upon enzymatic reactions, and these are influenced by a number of factors of which water content of the protoplast is an important one.

The decrease in water content of the leaves has a marked influence also on their amylolytic activity. In Table XVI are shown the results of half-leaf experiments in which one set of half leaves was kept with the petioles in water while in another set no water was supplied. The

amylolytic activity of the latter showed an increase of about 50 per cent at the end of the experiment as compared with the original condition of the leaves. In the first experiment, shown in Table XVI the leaves remained in the dark at 22° for 18 hours, in the second one at 23° for 16 hours. During this time no significant changes in the hydrogen ion concentration of the expressed saps were observed in any of the sets of leaves. It seems highly probable, therefore, that the increase in starch dissolution under wilting conditions is associated with the higher amylolytic activity which the leaves attain under these conditions.

TABLE XVI

THE EFFECT OF DECREASE IN WATER CONTENT ON THE AMYLOLYTIC ACTIVITY OF SUNFLOWER LEAVES KEPT IN THE DARK FOR 18 HOURS IN THE FIRST CASE AND FOR 16 HOURS IN THE SECOND

Conditions	Original		Final		Per Cent Change in Relative Activity
	Relative Activity	pH	Relative Activity	pH	
With H_2O, 22°........	100	6.83	95.2	6.90	− 4.8
Without H_2O, 22°......	100	6.85	144.8	6.87	+44.8
With H_2O, 23°........	100	6.98	103.0	7.05	+ 3.0
Without H_2O, 23°.......	100	6.97	156.5	7.00	+56.5

IV. EFFECT OF TEMPERATURE ON STARCH DISSOLUTION AND ON AMYLOLYTIC ACTIVITY OF LEAVES

As is the case with most enzyme preparations, those of amylase follow very closely the reaction rate-temperature rule, so that an increase of 10° results in approximately double the rate of reaction. This applies to amylase preparations at temperatures at which the inactivation of the enzyme does not come into play. Thus, the temperature coefficient, $k_{t+10} : k$, of the amylase reaction, below 30° has been reported to be very close to 2; Oppenheimer (1925, p. 153), Chrzaszez (1923), Ernström (1922).

A determination of the temperature coefficient of the amylolytic activity of ground fresh leaves was made in the following manner. The fresh sunflower leaf material was treated with toluene and ground as described under "Methods of Determining Amylolytic Activity" in Section II of this paper. A reaction mixture, containing ground fresh sunflower leaf material, was divided into five portions. One portion was kept at 0° in a bath of water and ice. The other four portions were

kept in thermostats at 5.0°, 15.0°, 25.0° and 30.0°. The "k" of each portion was determined.

Temp.	0°	5°	15°	25°	30°
k	.002386	.003488	.006425	.009728	.01036

The five points were plotted and a smooth curve was drawn through them. By interpolation, $k_{10}° = .00474$ and $k_{20}° = .00840$. $Q = k/k'$ was calculated for five temperature intervals:

Temp.	10/0	15/5	20/10	25/15	30/20
Q_{10}	1.987	1.842	1.772	1.514	1.233

In the experiments of Deleano (1911) with grape leaves by means of the iodine staining method it was found that the rate of starch disappearance was greater at 28° that at 19° and also greater at 19° than at 8°. That there are a number of factors besides the action of the amylase, which influence the rate of starch disappearance, such as the rate of respiration, hardly needs special emphasis. It is, however, of significance in view of the aforementioned facts concerning the influence of temperature on the rate of reaction of amylase, that the rate of starch dissolution in some leaves may proceed at the same rate, and even at a higher rate, at 5° than at 22°. Some examples of this effect are shown in Table XVII. Comparative determinations of starch dissolution at

TABLE XVII

THE EFFECT OF DIFFERENT TEMPERATURES ON THE STARCH DISSOLUTION OF LEAVES, BEFORE AND AFTER BEING KEPT AT 5° AND AT 22° FOR 18 HOURS IN AIR, SHOWN AS PER CENT OF STARCH OF THE DRIED LEAVES

Plant Material	Temp., ° C.	Original		Final		Per Cent Starch Decrease
		Per Cent Starch	Per Cent Dry Weight	Per Cent Starch	Per Cent Dry Weight	
Sunflower, mature	22°	4.15	21.49	1.36	18.90	67
Sunflower, mature	5°	4.31	20.63	0.80	19.80	81
Sunflower, young	22°	2.78	16.28	0.30	15.20	89
Sunflower, young	5°	2.83	16.26	0.39	15.70	86
Tobacco	22°	30.00	27.18	27.66	24.26	8
Tobacco	5°	28.40	26.96	25.42	25.86	10

22° and at 5° were made by means of the half leaf method, the leaves being kept in air at these temperatures for 18 hours.

These observations of the increased rate of starch dissolution in leaves at lower temperatures are in conformity with those of Lidforss (1896), who found that the photosynthetic tissue of evergreen leaves are starch-free during the winter.

The relative high rate of starch dissolution in leaves at 5° as compared with that taking place at 22° may in a measure find an explanation in the fact that the amylolytic activity appears to increase in leaves which have been kept at the lower temperature. Comparative experiments with two sets of half leaves kept in the dark, the one set kept at 22° and the other set at 5°, have shown that after 22–42 hours under these conditions the leaves kept at the lower temperature had a higher amylolytic activity than the ones kept at the higher temperature. During this period none of the sets of leaves showed significant changes in the hydrogen ion concentration of the expressed sap, so that the increase in activity cannot be ascribed to this. In Table XVIII are given some results of this type of experiment with leaves of tobacco and sunflower.

TABLE XVIII

EFFECT OF DIFFERENT TEMPERATURES ON THE AMYLOLYTIC ACTIVITY OF LEAVES KEPT IN THE DARK

Plant and Time	Temp., ° C.	Original		Final		Per Cent Change in Relative Activity
		Relative Activity	pH	Relative Activity	pH	
Sunflower, 18 hrs.....	22°	100	6.95	94.3	7.00	− 5.7
Sunflower, 18 hrs.....	5°	100	7.00	107.6	7.10	+ 7.6
Sunflower, 42 hrs.....	22°	100	6.90	94.5	6.95	− 5.5
Sunflower, 42 hrs.....	5°	100	6.92	103.4	7.00	+ 3.4
Tobacco, 18 hrs......	22°	100	6.05	57.4	5.93	−42.6
Tobacco, 18 hrs......	5°	100	6.32	126.8	6.32	+26.8
Tobacco, 42 hrs......	22°	100	6.10	89.3	5.97	−10.7
Tobacco, 42 hrs......	5°	100	6.12	134.7	6.12	+34.2

The increased amylolytic activity of leaves kept at low temperatures is not without interest in connection with the results of Manskaja and Schilina (1931), who reported an increase in the amylase activity of the wood of trees during the months of January and February when the average monthly temperatures were − 0.9° and − 2.5° and a decrease in activity with the higher temperatures in spring.

The final results of the decrease in water content of leaves and of their being exposed to what may be considered as subnormal temperatures is in general of the same nature so far as the starch is concerned. In both cases there takes place apparently an acceleration of the hydrolysis of starch with presumably an increase in the concentration

of the soluble sugars. This it may be assumed results in an increase of the osmotic pressure within the cells and a corresponding decrease in the freezing point of the cell fluids and of the vapor pressure thereof. These would tend to mitigate the influences of the environmental conditions to which the leaves are exposed. There is indication that this accelerated starch hydrolysis is conditioned by the amylolytic activity of the leaves, though the causes underlying this are still obscure.

V. Starch Dissolution in Killed Leaves

One of the striking facts of the amylolytic activity of leaves is that the dissolution of starch does not take place in leaves which have been killed, for example, by freezing or with toluene or chloroform. This was noted by older workers as Böhm (1883, p. 51) and by Brown and Morris (1893). From this observation Böhm concluded that the amount of enzyme present in the leaves at any one time must be exceedingly small and suggests that the enzyme is formed at about the rate it is used. Brown and Morris were apparently struck by this phenomenon and state (1893, p. 653): "Attempts made . . . to demonstrate the dissolution of leaf starch of any particular plant by the diastase of its own leaf were only partially successful." There is, however, no doubt that killed leaves contain amylase, and it has been shown in this paper that when such leaves are finely ground and the mass added to soluble starch this is hydrolyzed to reducing sugars. Our own experiments have shown that in leaves containing starch this does not disappear after the leaves have been frozen or treated with toluene, chloroform, potassium cyanide, or by any agent which does not destroy the enzyme, even though thereafter the killed leaves are kept under conditions favorable to amylase activity, *e.g.*, in solutions of a pH of 5.0 to 6.5 and temperatures of 35°, and, of course, under sterile conditions.

In general, therefore, it is evident that leaves which have been killed in such a manner that their protoplasmic activity has been arrested without destroying the enzymes do not show a depletion of starch. It is interesting that Brown and Morris (1893, p. 655) actually set out "to treat the leaf tissue by some method which, whilst exerting little or no inhibitory influence on any contained enzyme, would at the same time actually *kill* the protoplasm, and bring it into such a condition that the products of starch hydrolysis could readily pass outward when the conditions for such diffusion were made favorable." However, they concluded that the leaves "could never be induced to show a diminution in their contained starch within any time at all commensurate with that in

which complete dissolution takes place in the still living leaf after it has been plucked." And they finally conclude that (p. 658): "taking the whole of the evidence into consideration, we are compelled to admit that, at any rate, the first stages of the action on the starch granules are dependent upon the life of the protoplasm. . . ."

The results of our own experiments fully confirm the conclusions of Brown and Morris. Inasmuch as these are consistently of a negative character they are given here only in brief summary. Many of the observations are of a qualitative character, and they are largely based upon the detection of starch-iodide with the aid of a microscope, though quantitative experiments were also carried out in connection with the essential features of this aspect of the problem. In passing we wish to join in the warning of Steinhoff (1930) concerning the necessity of testing a portion of each leaf, or part thereof, used in starch dissolution or starch formation experiments for the presence of starch, for single leaves show a marked individual variation in starch content, as well as in the rate of starch formation and depletion. The importance of this cannot be overemphasized; it has already been stressed by Stanescu (1927), and doubtless many of the contradictory results which have appeared in this field can be attributed to an insufficient consideration of controls and to inadequate sampling.

The Retention of Starch by Killed Leaves. Young leaves of white clover show a high rate of starch formation in the light and a high rate of starch dissolution in the dark. The leaves of tobacco, although they accumulate very large amounts of starch, accomplish this relatively slowly and show a low rate of starch dissolution. In these respects the leaves of sunflower are about midway between clover and tobacco.

The leaves of clover and sections of tobacco and sunflower leaves were treated with toluene or chloroform and placed in phosphate solution of a pH of 5.5. The stoppered flasks containing the leaves were kept at different temperatures ranging from 4.5° to 35° and examined periodically for starch. To prevent the growth of moulds, toluene or chloroform was added to the flasks. In no case was the disappearance of starch observed, although some experiments were continued for 143 days. Solutions varying in pH from 2.5 to 7.5, in whole units, were also used without different results. Nor did depletion of the starch occur when the leaves were on a silver gauze in a moist chamber. From the quantitative determinations of the starch content of leaves treated with chloroform it is also evident that such leaves do not lose starch. The same results were obtained with various species of leaves killed with toluene and kept in an atmosphere of oxygen or of irrespirable gases.

When leaves were killed with toluene and finely ground immediately, detectable starch dissolution occurred only in those species having very high amylolytic activity, as clover; in leaves of sunflower and tobacco no appreciable starch dissolution occurred even after four days at 30°.

In the case of leaves which have been killed by freezing, no indication of any starch dissolution was ever obtained. However, such leaves could be kept for only a few days without infection. If antiseptics such as chloroform are added, there is, of course, no way of telling whether the non-depletion of starch is due to the antiseptics or to the freezing. Similarly, dried leaves which were again moistened showed no starch depletion.

The Amylolytic Activity of Killed Leaves. That the non-dissolution of starch in killed leaves was not due to a destruction of the enzyme becomes apparent from the fact that leaves treated with toluene or chloroform can be shown still to contain amylase after relatively long periods of time, Fig. II.

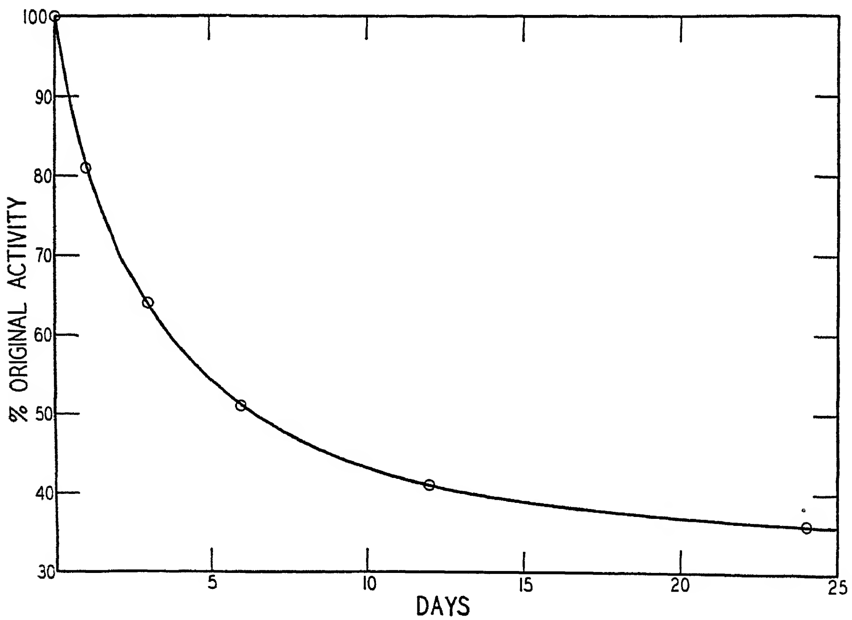


Fig. II. Decrease in amylolytic activity of sunflower leaf material after standing with toluene.

Protected Starch Granules. It seemed conceivable that in the process of killing the leaves, either through freezing or by means of chloroform, etc., the colloidal structure of the protoplasm was altered and resulted in the deposition of some substance around the starch

granule making it impossible for the amylase to gain access to the starch. From what little is known of the structure of the chloroplast it would appear that the chlorophyll and the carotenoids are in intimate contact with the starch-containing chloroplasts. The process of killing may affect this structure in such a manner that the starch granules become covered with a film or layer of oil or material through which the amylase can not diffuse.

Accordingly leaves which had been killed by freezing, drying in vacuum, or with toluene or chloroform were extracted with various solvents at temperatures from 15° to 60° and placed in phosphate solution of various hydrogen ion concentrations, at different temperatures, and examined periodically for their starch content. The extractions were made for various lengths of time with wet and dry leaf material. The solvents removed varying amounts of pigments, fats and other materials from the leaves. There were used: methanol, ethanol, acetone, petroleum ether, dioxane and xylene. With none of the leaf material treated in this manner was any evidence obtained of the dissolution of the starch.

It also seemed conceivable that during the killing the starch granules had become imbedded in the protein of the protoplasm and had thus become inaccessible to the amylase, as has been demonstrated for amylase preparation by Oparin, Manskaja and Glasunow (1934). Leaves which had been killed by different means were placed in solutions of pepsin and of papain of hydrogen ion concentration ranging from pH 2.5 to 7.5 and periodically tested for starch. Not a single combination of conditions was found which resulted in the dissolution of the starch.

Similarly, it seemed possible that in the killing of the leaves by these means the amylase present therein had become bound or inactivated or that, on account of the absence of protoplasmic streaming, the contained amylase did not suffice for the hydrolysis of the starch present. A series of killed leaf preparations were placed in solutions of takadiastase and of malt amylase with phosphate buffer solutions and tested periodically for starch for several months. In no case were the starch granules of the chloroplasts dissolved.

Pre-treatment. It is a familiar fact that starch granules, at least those obtained from seeds and storage organs, show an initial resistance to the action of amylase and that consequently the starch is far more easily hydrolyzed after the starch grain has been injured or ruptured mechanically or through the action of heat. The same may apply to the starch granules of the chloroplasts and it seemed possible that the

dissolution of the chloroplast starch was dependent upon some primary action before the amylolytic action could proceed, and that such action was accomplished by some other element of the protoplasm.

Leaves of clover, containing an abundance of starch, were placed in the dark for a preliminary period of from three to fifteen hours at 15° to 20°. After such periods in the dark distinct evidences of the beginning of starch dissolution were detectable. The leaves were then treated with toluene, placed in phosphate solution of pH 5.5 and examined periodically. No evidences of further dissolution of the starch could be detected even after sixty days. Similarly only negative results were obtained with clover leaves which had been given a pretreatment by heating at 52° for an hour and at 60° for 15 minutes.

Amino Acids. That amino acids have an accelerating influence on amylase activity has been clearly demonstrated by Sherman and his colaborators (1923). The addition of glycin, alanin and of pepton to various preparations of killed leaves in phosphate solution produced no perceptible results in bringing about the dissolution of the starch in the chloroplasts.

Effect of Wilting. It has been repeatedly demonstrated that with the loss of water the starch content of leaves diminishes rapidly, there being an increase in the rate of amylolysis during wilting of the leaves. If, however, the leaves are first frozen or treated with chloroform or toluene and then allowed to dry, no loss of the contained starch was ever observed; nor do leaves which have been thus dried, when again moistened, undergo starch dissolution.

It must be concluded that starch dissolution does not occur in leaves which have been killed by drying, freezing or treatment with chloroform or toluene, although such treatment is of itself not destructive to the amylase or to the starch contained in the leaves. There is no doubt, however, that such treatment results in marked changes in the finer structure of the cell and results in alterations of the colloidal phase relationships of the components of the chloroplasts and of the surrounding protoplasm.

For helpful criticism of various aspects of this work we are indebted to Dr. J. H. C. Smith and to Dr. H. H. Strain.

REFERENCES

Bailey, E. M. 1905. Studies on the Banana. I. *Jour. Biol. Chem.*, **1**, 355–361.

Böhm, J. 1883. Über Stärkebildung aus Zucker. *Bot. Zeitung*, **41**, 33–38; 49–54.

Brown, H. T., and Morris, G. H. 1893. A Contribution to the Chemistry and Physiology of Foliage Leaves. *Jour. Chem. Soc.*, **63**, 604–683.

Chrzaszez, T. 1923. Einfluss der Temperatur auf die Wirkung der Amylase. Biochem. Zeit., **142**, 417–439.

COLLORIO, H. M. 1928. Untersuchungen über die Beziehungen zwischen der Wasserabgabe der Pflanzen und ihrer Atmungsgrösse. *Planta, Arch. wiss. Bot.*, 5, 1–27.

DELEANO, N. T. 1911. Über die Ableitung der Assimilate durch die intakten, die chloroformierten und die plasmolysierten Blattstiele der Laubblätter. Jahr. wiss. Bot., 49, 129–186.

DETMER, W. 1883. Über die Entstehung Stärkeumbildender Fermente in den Zellen höherer Pflanzen. Bot. Zeitung, 41, 601–606.

EISENBERG, A. 1907. Beitrage zur Kenntnis der Entstehungsbedingungen diastatischer Enzyme in höheren Pflanzen. *Flora*, 97, 347–374.

ERNSTRÖM, E. 1922. Über den Temperaturkoeffizienten der Stärkespaltung und die Thermostabilität der Malzamylase und des Ptyalins. *Zeit. Physiol. Chem.*, 119, 190–263.

EULER, H., AND JOSEPSON, K. 1923. Bezeichnung der Aktivität und Affinität von Enzymen. *Ber. deutsch. chem. Ges.*, 56, 1749–1761.

EULER, H., AND SVANBERG, O. 1921. Über die Characterisierung von Amylaselösungen. *Zeit. physiol. Chem.*, 112, 193–230.

FIFE, J. M., AND FRAMPTON, V. L. 1935. The Effect of Carbon Dioxide upon the pH and Certain Nitrogen Fractions of the Sugar Beet Plant. *Jour. Biol. Chem.*, 109, 643–655.

GRÜNBERG, G. 1932. Über die Ursachen des Zelltodes in Anaerobiose. *Planta, Arch. wiss. Bot.*, 16, 433–466.

ILJIN, W. S. 1923. Einfluss des Welkens auf die Atmung der Pflanzen. *Flora*, 116, 379–403.

——. 1930. Der Einfluss des Welkens auf dem Ab und Aufbau der Stärke in der Pflanze. *Planta, Arch. wiss. Bot.*, 10, 170–184.

KERN, H. 1928. Die diastatische und peroxydatische Wirksamkeit in der Pflanze nach deren Beeinflussung durch Aussenfaktoren. *Zeit. f. Bot.*, 21, 193–252.

KISSELEW, N. W. 1927. Zur Frage des Stärkeabbaues beim Welken der Blätter. *Planta, Arch. wiss. Bot.*, 4, 606–616.

LIDFORSS, B. 1896. Zur Physiologie und Biologie der wintergrünen Flora. *Bot. Zentralbl.*, 68, 33–44.

MANASKAJA, S., AND SCHILINA, M. 1931. Der Fermentgehalt in den Baumpflanzen während der Winterruhe. *Biochem. Zeit.*, 240, 276–285.

MOLISCH, H. 1922. Über den Einfluss der Transpiration auf das Verschwinden der Stärke in den Blättern. *Ber. deutsch. bot. Ges.*, 39, 339–344.

NEGER, F. W. 1919. Die Blattrollkrankheit der Kartoffel. *Zeit. f. Pflanzenkrank.*, 29, 27–48.

OPARIN, A., MANSKAJA, S., AND GLASUNOW, I. 1934. Die Inaktivierung der Amylase durch Adsorption an Eiweissniederschlägen. *Biochem. Zeit.*, 272, 317–323.

OPPENHEIMER, C. 1925. Die Fermente und Ihre Wirkungen. Leipzig, 5th Edition, Vol. I.

PAECH, K. 1935. Über die Regulation des Eiweissumsatzes und über den Zustand der proteolytischen Fermente in den Pflanzen. *Planta, Arch. wiss. Bot.*, 24, 78–129.

PALLADIN, W., AND POPOFF, H. 1922. Über die Entstehung der Amylase und Maltase in den Pflanzen. *Biochem. Zeit.*, 128, 487–494.

PHILLIS, E., AND MASON, T. G. 1937. On the Effects of Light and of Oxygen on the Uptake of Sugar by the Foliage Leaf. *Ann. Bot.*, N. S., 1, 231–237.

RYWOSCH, S. 1908. Zur Stoffwanderung im Chlorophyllgewebe. *Bot. Zeitung*, 66, 121–130.

SCHROEDER, H., AND HORN, T. 1922. Das gegenseitige Mengenverhältnis der Kohlenhydrate im Laubblatt in seiner Abhängigkeit vom Wassergehalt. *Biochem. Zeit.*, 130, 165–198.

SHERMAN, H. C., AND WALKER, F. 1923. Effect of Amino Acids in Retarding the Hydrolytic Decomposition of an Enzyme (Pancreatic Amylase). *Jour. Amer. Chem. Soc.*, 45, 1960–1964.

SJÖBERG, K. 1922. Beiträge zur Kenntnis der Amylase in Pflanzen. I. Über die Bildung und das Verhalten der Amylase in lebenden Pflanzen. *Biochem. Zeit.*, 133, 218–293.

SPOEHR, H. A. 1919. The Carbohydrate Economy of Cacti. Pub. No. 287, Carnegie Inst. of Washington.

—— AND MILNER, H. W. 1936. The Starches isolated from Plant Material by the Freezing Method. *Jour. Biol. Chem.*, 116, 493–502.

STANESCU, P. P. 1927. Les Variations Quantitatives de Produits de la Photosynthese dans les Feuilles des Plantes vertes pendant un Jour (Vingt-Quatre Heures). Dissertation. Bucharest.

Steinhoff, E. 1930. Über den Einfluss von Salzen auf den Stärkeabbau in Blättern einiger Land- und Wasserpflanzen. *Planta, Arch. wiss. Bot.*, **11**, 207–242.
Thornton, N. C. 1933. The Influence of Carbon Dioxide on the Acidity of Plant Tissue. *Contrib. Boyce Thompson Inst.*, **5**, 403–418.
Tollenaar, D. 1925. Omzettingen van Koolhydraten in het Blad van Nicotiana Tabacum L. Dissertation. H. Veeman and Zonen, Wageninen.
Winkler, H. 1898. Untersuchungen über die Stärkebildung in den verschiedenartigen Chromatophoren. *Jahr. wiss. Bot.*, **32**, 525–556.
Wortman, J. 1890. Über den Nachweis, das Vorkommen und die Bedeutung des diastatischen Enzymes in den Pflanzen. *Bot. Zeitung*, **48**, 581–594; 632–654.

THE PRESERVATION AND ABUSES OF MOTION PICTURE FILM BY SCIENTIFIC INSTITUTIONS [1]

E. R. FENIMORE JOHNSON

(Communicated by Roland S. Morris)

Abstract

The writer's findings as to what the present status of motion picutre film is in scientific institutions are given, and there is speculation as to its future.

There is a description of the best means of handling film in and about institutions and of the best storage conditions.

Facts are stated which enable the reader to judge for himself the desirability of the various kinds of film as material for permanent records.

The writer recommends 16 mm. cellulose acetate base film as the best at present available for this purpose.

Steps other than correct storage which must be taken in order to assure permanent retention of valuable motion pictures are described. There is a list of necessary equipment, and the more important literature pertaining to the evaluation of motion picture film for permanent records and its proper care and storage is cited.

Introduction

Up to the present year scientific institutions in general have not regarded motion pictures as an important means of recording their work for posterity. The pictures taken by both their amateur and professional staffs have been, almost without exception, of poor quality photographically and editorially. It remained for the teachers, the surgeons and the popular lecturers to demonstrate the great value of supplementing written and still photographic records with motion pictures.

This state of affairs does not indicate an inferior ability on the part of the pure scientist. It happened because the greater desire of the other groups to disseminate knowledge impelled them to call upon technically trained motion picture men for assistance. The teaching films have demonstrated beyond all rational doubt that it is possible to record on a few feet of motion picture film facts which either could not be recorded at all or which would require a hundred times as much effort to record by any other means.

[1] *Editor's note:* In connection with the microfilm outfit which has been established in the Library and also as a supplement to the consideration of newer methods of publishing and recording research work, which were considered at the mid-winter meeting of 1938, the following paper on the scientific uses of motion picture film has been offered to the American Philosophical Society by Mr. E. R. Fenimore Johnson, whose interest in the Society has been shown in his authorization of the establishment of the Eldridge Reeves Johnson Fund for the promotion of useful knowledge.

The arrival of accurate and comprehensive educational films, the widespread popularity of amateur cinematography and the mass production of cinematographic equipment, with resulting price reductions, make it safe to predict that the purely scientific institutions will soon be forced to meet the problem of the proper use and storage of motion picture films.

The collections of films now in their possession contain an appreciable number of exceedingly valuable scenes. These valuable increments have generally been lost track of through the separate editing of each batch of film with an eye only to portraying the current work of some individual or party of students.

The customary motion picture of an expedition is a chronological portrayal of its methods of travel and the hardships of its members, but in nearly every one there are a few scenes which show something of value to science; a native custom, a method of manufacture, something of geographical interest or the natural activities of an animal, and in some cases irreplaceable records of departed people, animals and customs. Now suppose that after fifty expeditions a naturalist were to examine his films en masse. He would probably pick out all the pictures of some one animal. He would show its development from embryo to adult, its method of locomotion, of feeding and defense. He would compare the type that developed in Asia with that in America. Perhaps he would show environmental scenes to explain their differences.

When he had achieved a smooth sequence of scenes he might add animated drawings, charts and diagrams of his own conception. Then finally sounds and verbal comments or even a complete discussion of his subject keyed to his illustrative scenes. He would then have created a truly scientific record of the animal in question, and what is almost of equal importance, he could quickly and easily transmit the record to others.

Now consider his chance of performing this service by personally causing the necessary scenes to be photographed. He would have to go on innumerable trips to a great many places or arrange for others to do so. The proper portrayal of one animal would take a lifetime.

This paper does not attempt to discuss the advantages, disadvantages and correct methods of making scientific records by means of motion picture film. They constitute an extensive problem which should be discussed in a separate paper. Assuming that the use of motion picture film is advantageous in the making of scientific records, and that the use of motion picture film is soon to become extensive in scientific institutions, and that it is the moral obligation of these institu-

tions to rescue such of it as they have on hand, this paper is restricted to a description of the best means of handling film in and about the institutions and of the best storage conditions.

In many cases when the film has been permanently retained by the institutions it has been stored incorrectly, sometimes, in the case of nitrate films, introducing hazards to human life, not to mention the certain destruction of all types within a very few years by heat and dryness.

In many more cases the patrons and the scientists have carried off both positive and negative film for their personal purposes and never returned them.

Pertinent Facts

Nitrate Film

Because serious deterioration of nitrate film can occur within four or five days after development it must be chemically tested to insure that it has been properly fixed and washed, and that it will have a normal life under correct storage conditions. It is also worthy of note that even when stored under the most ideal conditions nitrate film will not last indefinitely and when under extremely undesirable conditions, even though properly washed and fixed, it may become permanently brittle and undergo serious chemical deterioration within ten days. After being stored under ideal long time storage conditions nitrate film must be tempered to avoid condensation upon removal.

As nitrate film is believed capable of spontaneous combustion, storage requires not only the protection of the surroundings in the event of a fire within a vault, but the proper isolation of film within the vault, so that only the minimum amount of film will be destroyed. Fire in nitrate film can only be extinguished by reducing its temperature with water as it has a low ignition temperature combined with a high rate of combustion, and is capable of decomposition with little or no air supply. It is therefore evident that suffocation methods are of no value. This type of film can explode with terrific force accompanied by the generation of poisonous gases. An example of the danger from nitrate film is the horrible film fire in Cleveland. Much of the film was not burning but was heated and gave off poisonous gases. That such gases will spread over large areas is illustrated by the case of the man who drove through some of the smoke and whose car crashed into a pole less than a block away. The man was dead when removed from his car, due to his having inhaled some of the gas.

Gas resulting from burning nitrate film is not so bad as gas resulting from the heating of nitrate film, as illustrated by the number of booth

fires in which the projectionist apparently suffers no harm. Film that is decomposing rapidly but not burning gives off oxides of nitrogen, and when these gases come in contact with moisture in the lungs nitric acid is produced, which is poisonous and very irritating. It is, however, only in recent years that these conditions have been fully appreciated. There is no telling how many deaths have been due to projectionists having inhaled the fumes, because their deaths have been blamed on heart failure and various other things.

Prolonged storage of any quantity of nitrate film must be in a vault constructed in accordance with the National Board of Fire Underwriters' specifications. In this connection it is noted that the storage of nitrate film negative in basements is prohibited. Nitrate film storage vaults, even without air conditioning, are exceedingly complicated, very expensive and the entire science of film storage is in a state of flux. A few of their more salient features are fireproof walls, skeletonized shelving, vents equipped with traps having openings outside of the building at least fifty feet away from any other opening, sprinkler systems and special electrical lighting equipment. When not stored in valuts or being directly worked upon, nitrate films must be stored in vented cabinets having a strictly limited capacity. It is important to note that in the storage of nitrate film the vaults or cabinets should be kept at a controlled humidity not above fifty per cent, and that the temperature should always be kept below fifty degrees Fahrenheit.[1] Because of the air-conditioning these vaults are very expensive. The construction and arrangement of rooms in which nitrate film is to be worked on are subject, also, to extensive, complicated and therefore expensive regulations. A few of these regulations state that the rooms must have two exits, special vents, special spacing of workers, specially designed tables, racks and illuminating equipment, in addition to a host of rules and regulations pertaining to the operation of the work in these rooms. When nitrate film is to be stored in outside buildings it is necessary that these buildings be at least one hundred feet from any other building, and meet with many complicated and expensive requirements. Film cans should be ventilated. For permanent storage of nitrate film a specially constructed and expensive can is required. For ordinary handling and shipment of small quantities the common film can is sufficiently well ventilated.

The life of nitrate film is not definitely known; but even when correctly stored, it is thought not to retain the best of condition more than thirty or forty years. The danger of shrinkage possibly even shortens this estimate. Deterioration occurring in nitrate film is not necessarily

[1] C. G. Weber and J. R. Hill, December, 1936.

visible. There is a gradual shrinkage of film due to the loss of volatile materials which cannot be guarded against by any means other than duplication at stated intervals. In addition to the expense of storing it there is a strong possibility that nitrate film is on its way out. Therefore, in the case of an institution having a small supply of nitrate print or nitrate negative on hand it is recommended that the prints be disposed of, new safety prints bought and the nitrate negative be placed in storage with a commercial storage company.

Nitrate film is used by theatrical picture producers because when used promptly after development it is less expensive and better able to withstand the heat and mechanical wear of theatrical projection than is acetate film.

Safety (Acetate) Film

The ignition temperature of safety film is between seven hundred and eight hundred degrees Fahrenheit as compared with about three hundred degrees Fahrenheit for nitrate film, the fire hazard being classed as somewhat less than that of common news print paper in the same form and quantity. In contrast with nitrate film the decomposition of safety film once started does not continue except where there is an external source of heat, but if ignited will continue to burn as long as the supply of air is sufficient to support combustion freely. If the supply of air is restricted the combustion will cease, thereby demonstrating another advantage over nitrate film, which is capable of combustion with little or no air supply. Under the law, if safety film is handled or stored in conjunction with nitrate film it must be treated as though it were nitrate film in so far as fire hazard is concerned.

In quantities up to about five hundred pounds safety film can be safely stored without special precautions against fire, but where very large quantities are to be stored it is advisable to provide cabinets of fireproof construction. As to temperature and humidity the best conditions are the same as those recommended for the preservation of books in libraries, namely; relative humidity of fifty per cent and temperature of seventy to eighty degrees Fahrenheit. It is essential that the relative humidity should not drop below fifty per cent, as this value is high enough to prevent brittleness, yet does not introduce the condensation problem that might be encountered at higher humidities. Extreme variations of moisture content affect the flexibility and cause dimensional changes. It appears that some air conditioning precautions are essential for the storage of safety film as tests indicate extreme brittleness at a relative humidity of about fifteen per cent and below, and as fifteen per

cent is within the range of humidity encountered during the winter season in heated rooms with no humidity control.

Even though safety film has been allowed to become brittle through excessive drying this condition does not cause irrevocable loss, as the flexibility can be restored with the restoration of moisture. Safety film can, however, deteriorate rapidly due to residual hypo if it has not been properly washed. The presence of residual hypo can be detected by a method recommended by Crabtree and Ross.[1] Some foreign safety films are manufactured with small quantities of cellulose nitrate, which is undesirable and which can be detected by means of a diphenylamine indicator. It is also worthy of note that the life of safety film can be prolonged by a vapor film treatment at moderate cost.

For storage of safety film over a long period of time in the average institution, a large home type or store type electric ice machine, adjusted to a temperature of about seventy degrees and supplied with humidifying pans, should be adequate. This type of equipment is comparatively inexpensive, and the units can be duplicated as they are needed. These safety film storage cabinets should, of course, bear signs prohibiting the storage of nitrate film in them. The reason why a temperature of about seventy degrees instead of lower temperatures is recommended is to avoid any risk of condensation on the film when it is removed. There seems to be little need for using a lower temperature for the storage of safety film than seventy degrees as it has about as long a life at that temperature as it would have at a lower temperature of about forty degrees, where the condensation problem would be serious.

Safety film is a very stable material, being comparable to permanent record paper if properly made, and if care is taken to prevent it from drying out excessively.

Common to Both Types

When positive films are in constant use they require occasional replacement, therefore the importance of carefully preserving negative films of important records as a source for replacement cannot be minimized. Films in constant use should be kept clean, not only because extraneous material on the film affects legibility, but because dust particles cause scratching of the film when in use. After they are used in the projection machine they should be exposed to humidified air in such a way that the air has access to all parts of the film, to restore moisture that may have been lost, and they should not be re-used until moisture equilibrium has been obtained.

[1] *Jour. Soc. Motion Picture Engrs.*, 14, 419 (1930)

If a reel of film is tightly wound, even though stored in air conditioned chambers for six months, it will not recover the moisture lost during projection and should, therefore, be placed in a humidified cabinet or room after projection and before storage, in open reels without cans. All tightly wound rolls should be loosened and hung exposed so that the air will have free access to all parts of the film. Successive projections drive out the film moisture more rapidly than it can be regained while tightly wound, and even though infrequently used may thus become brittle although stored in the humidified atmosphere between projections. Thorough rehumidification can be accomplished in ten to thirty minutes by passing the film through a humid chamber, or by other methods such as moistening with a mixture of water and a water miscible volatile liquid. In libraries without conditioned air, satisfactory conditioning of films can be accomplished by means of humidification in small closed rooms or cabinets. Small closed rooms can be humidified with relatively inexpensive air conditioning units. Cabinets or small vaults can be humidified by using open vessels of the proper salt solution. A saturated solution of dichromate of sodium, because it gives a relative humidity of fifty-two per cent at sixty eight-degrees Fahrenheit, should prove very satisfactory for the purpose.[1] The solution should be exposed in shallow vessels with fans blowing across the surface to promote the circulation of air. Inasmuch as a considerable amount of water may be absorbed from the air an excess of dichromate should always be used to insure satisfaction.

In the storage of all kinds of film, the practice of some commercial film storage companies notwithstanding, it should always be remembered that they should be on edge; that is, with the rolls lying on their sides.

16 mm. Film Versus 35 mm. Film

In the beginning a discussion of the relative qualities of 16 mm. film and 35 mm. film the prediction is made that in the future all projection prints for other than theatrical use will be entirely 16 mm. due to the great economies and the satisfactory quality for lecture hall purposes; but a large percentage of negative exposed will continue to be 35 mm. The 16 mm. negative, while customarily reversed during development into a positive, may, by special instructions to the company, be developed as a negative, thus giving it the same flexibility of editing and duplication as the 35 mm. 16 mm. film is all safety film. Its small size is a great advantage in storage, and the equipment used in the taking of the 16 mm. pictures is much more portable than that of the 35 mm. On

[1] C. G. Weber and J. R. Hill, November, 1936.

a short throw, where the distance between the projector and the screen is not very great, the quality of the 16 mm. projection is substantially equal to that of the 35 mm.

The producer of a motion picture must determine in advance whether he will ever want to use it for theatrical purposes because the 35 mm. film is the only one that can be projected over such distances and spread out to such screen sizes as are common to theaters without showing defects, and because negatives of 16 mm. width can seldom, if ever, be enlarged to 35 mm. with sufficient quality to be acceptable to theaters. On the other hand, 35 mm. film can be reduced to 16 mm. size without loss of quality.

8 mm. Film

In regard to museum use 8 mm. film is hardly worthy of consideration as it is not of satisfactory quality for lecture hall purposes, and although a rapid advance is being made in this size film it does not seem probable that it ever will satisfy the needs of scientific institutions.

Color Motion Picture Film

As yet little is known about the lasting qualities of color motion picture film in storage, and there is nothing to indicate which type of color film will achieve general usage.

Nevertheless, color is such an important part of a scientific record that the deferment of the use of color motion picture film by scientific institutions cannot be recommended.

Some—possibly all—color motion picture film is high unstable in the presence of percentages of humidity which are not too high for black and white film.

It is regretted that the correct temperature and humidity for the preservation of color film cannot be given at this time, but as most, if not all, color film is upon a safety—that is, acetate base—it is probable that the best storage life will be attained between the temperatures of 60 and 70 degrees Fahrenheit and relative humidity between 40 and 50 per cent.

Most color film can be treated by the vapor processes to increase its life and may be cleaned with the cleaning fluid commonly used upon black and white film. However, the manufacturer of the color film which is to be cleaned or treated should be consulted before such steps are taken.

The saving grace of present day color film is that most of it can be developed in negative form and that black and white prints can be made from it, which are reasonably comparable though not absolutely equal

in quality to prints which have been made from ordinary black and white negative.

Therefore, in the case of color pictures which are expected to have permanent value their preservation should be assured by at least the making of a black and white print from the color negative, and where the expense is warranted a black and white duplicate negative; in both instances using, of course, acetate (safety) film.

It may be possible to make black and white negatives directly from color print.

Standard Film Speeds

The change of standard from 16 to 24 frames exposed per second which occurred when the industry switched from silent to sound motion pictures created a serious difficulty for anyone wishing to combine pictures taken before the change with those taken subsequently.

Pictures exposed at the rate of 16 frames per second and projected at 24 frames seldom portray action in a pleasing manner, and still more seldom do they constitute a record having an accuracy worthy to be called scientific.

While effective in the case of views of inactive subjects, the corrective device of double printing of each frame is no more than a means of extending the time during which a given scene remains before an audience. The speed of the subject's actions is more distorted than when double printing is not used. Therefore, it is most important that the practice of taking pictures at 16 frames per second, which is still indulged in on the grounds of footage economy by some scientific students, be discontinued.

This applies to both the 16 and the 35 mm. film sizes.

Copyright

The right to copyright is very easily lost.

Information on how to proceed to obtain copyright protection can be obtained from the Library of Congress.

Full information should be obtained before any public showing of any nature if copyright protection is contemplated.

Copyright protection for motion picture film is quite expensive because three copies correctly titled and with scrupulously correct notices must be filed with the Library of Congress before copyright is granted.

Routine Recommended for Acceptance, Use and Storage

Much has been said of the necessity for testing the original negative and the prints for chemical residue in order to make sure that they have been properly fixed and washed.

Fixing and washing is a matter applying to each individual roll, and the testing of one roll does not prove that any other roll is going to be satisfactory. It is also quite possible for portions of any one roll to be improperly fixed and washed.

With these difficulties in mind it is suggested that the whole question of testing for fixing and washing be left to the care of a commercial developing laboratory.

In attempting to describe the ideal procedure for a scientific institution, let us assume that it has persuaded the patron or research worker to turn over the negative as they expose it. Upon reception it should be immediately placed in the hands of a commercial developing laboratory, the staff of which will develop all of it, place it in storage and await further instructions from the institution.

When the time arrives to begin the work of editing and splicing a picture from the material, the commercial laboratory should be told to make a print of all the footage which appears to be properly exposed and to have no mechanical defects. Upon receiving these prints the person undertaking the editing should take the field notes of the expedition and identify each scene, write out a small card known as a "scene card" which describes the scene, gives the footage, and identifies by means of roll numbers and marginal numbers the exact part of the negative from which each scene was printed.

Having examined in detail and described each scene, the editor should proceed to place the scene cards in the order in which he expects to splice the film.

Taking next the print that has come from the laboratory, and which is hereafter known as the "Work Print," he should then cut the scenes apart and splice them together in the order planned.

The story having been told as clearly as possible by means of the scenes in hand, and after inspection and trimming down of the flashes and other defects inevitably found in the "work print," only then should an order be placed with the film laboratory for titles, maps, animated drawings and other things required for the finished picture. Upon receipt these are also spliced into the "work print." In many instances the "work print" is all that it is ever worthwhile to make from negatives depicting scientific subjects, and it serves well enough the additional function of "projection print."

The use of the work print for projection should be discouraged, however, if the completed story is expected to have any permanent value. Assuming that the story of the particular piece of scientific work in question as edited for current viewing has been well told and is in a

form which will be of wide or permanent interest, thus requiring to be shown many times, the "work print" should then be sent to the cutting laboratory together with a scene list, which should show the scene sequence, the footage, and give detailed instructions as to which part should be made a little brighter and which part a little darker. From the guidance of the "work print" and the written instructions the laboratory will proceed to make up a "projection print."

It may be urged by the developing laboratory and it is the natural tendency of film editors to cut apart the original negative and splice it together into the story currently desired. This is a bad practice for anyone except amateur home movie makers and the big theatrical producers and should absolutely never be indulged in by anyone who aspires to do scientific work. The reason for this will be made very clear later.

If given accurate directions, a commercial film laboratory can make a print by picking out the scenes one at a time from the "original negative" rolls, provided the rolls are not more than 200 feet in length. Although this is slightly more expensive than printing from a negative cut and spliced together, in the case of museums where the maximum demand for any given arrangement of the scene will probably not be for more than three or four prints over a period of many years, the increase in cost is small.

If it is expected that new identical prints will be needed when the initial "projection prints" wear out, then before projecting these prints more than is necessary to ascertain if they are in good order, it is advisable to have made what is known as a "duplicate negative." With a "duplicate negative" on file it is always possible to duplicate the original picture quickly and easily and yet keep the "original negative" uncut. This "duplicate negative" should be safety film.

In editing pictures of scientific subjects it is frequently necessary to borrow sections from old pictures and splice them together in such a way as to tell a new story. The reasons why it is so important to keep the "original negative" uncut and intact in the rolls just as they were exposed are:

First, the field or laboratory notes can seldom be identified with the scenes to which they pertain by other means than by roll numbers and scene sequence. Once the identity of a roll is lost and the original sequence of the scenes is lost the task of correlating notes and scenes can rarely be accomplished even by the person who exposed the film.

Secondly, in editing a picture for current interest numerous good scenes which do not suit the continuity of the story planned by the editor are left unused, and yet these currently unused scenes are often most

valuable to persons who perhaps years later edit pictures to cover possibly a complete field of scientific work or a special section of a field long after the detailed experiences of a given expedition or group of research workers are no longer of general interest.

It is therefore recommended that the negative once placed in the hands of the commercial laboratory never be removed therefrom until such time as it is desired to place it in permanent storage.

In a case where the negative is nitrate if the museum has an air conditioned vault meeting underwriter's requirements it may safely store in its own vault; otherwise, permanent storage should be purchased from some commercial company making a practice of selling storage space. The use of nitrate film by scientists should be vigorously discouraged, and nitrate negatives of any value to posterity should be duplicated in acetate film.

Now, in addition to the problems pertaining to the "original negative" and the manufacturing of "projection prints," there are points to be remembered if maximum economy and satisfactory quality are to be obtained when prints are projected. Assuming that a print has just been projected and has been returned to the care of the institution for servicing and re-storage, the print should be inspected for breaks or partial breaks. This is generally done manually, the film being wound across from one reel to another and the operator allowing it to run over his finger so that any partial breaks or tears are felt as they pass. When partial breaks or tears are felt, the frame is cut out and the film re-spliced. Although this process gradually shortens the print, it is of no importance until it becomes quite extensive.

After being repaired the print should be run through a machine which both cleans and humidifies it, and it should then be immediately replaced in the storage cabinet, since if it is allowed to lie about unprotected it may acquire dust particles which will scratch it upon its next projection or lose moisture to such an extent that breakage will become frequent. Given proper care the life of a film should be at least 150 projections, often far more, but if projected while dry or dusty, the film may be substantially ruined by one or two projections.

Equipment Recommended and Space Required

General Equipment for Film Work Room

The usual equipment of editing and cutting rooms is as follows:

Table space with portable chairs.

Rewinding spindles, hand driven.

Splicers; that is, small pieces of equipment to hold the film in place while it is cemented together, and which also act as cutting tools.

An editing machine, preferably with sound. This is a machine which instead of projecting the film on a curtain or screen, shows it under a magnifying glass. On this machine the film can run forward or backward in order that it may be inspected a number of times without rethreading.

A cleaning and humidifying machine.

Temporary storage cabinets. In the case of safety film the long time or permanent storage cabinets can be kept in the work room.

Large waste baskets for clippings.

A board full of pegs or other devices for filing small rolls of one or two scenes, pending the time that they are spliced together.

Some small compartmented boxes of correct size to hold scene cards.

A typewriter.

Extra film cans and reels.

A supply of film softener.

Hand-type picture magnifiers or inspection glasses.

A supply of film carrying cases in various sizes.

Some wooden spools and discs for winding short lengths of film.

It is not advisable for the average institution to equip itself for the making of titles or animated drawings, or to combine any function of the studio or commercial film laboratory with the editing and splicing room. It is advised, however, that even though no nitrate film is contemplated being used therein, the editing and splicing room be fireproof and fulfill the requirements of the fire underwriters in regard to the positions of tables, the covering of radiators and the use of safety type electrical equipment.

Conclusion

In conclusion it is recommended to museums not already equipped for editing, reconditioning and storage of motion picture film, that they set aside and equip a room for such purposes; that the room selected shall be at least as fireproof as libraries; that unless the motion pictures they expect to produce are to be shown in theatres, their purchases of cutting, editing and storage equipment be confined to the 16 mm. size; that they urge both students and patrons to leave their negatives and correlated notes permanently with the institution.

Where 35 mm. nitrate negative is used, a duplicate negative of acetate film should be made and filed.

When the pictures are taken upon 16 mm. film, it should not be reversed to positive, but should be developed as a negative and prints made therefrom in the usual manner.

The standard film velocity of 24 frames per second should be adhered to. Reservoirs of negative film, indexed and classified by subject should be created. Their purpose is not so much to insure the possibility of duplicating pictures as they are edited for exhibition while interest in a particular student's or expedition's work remains alive, but to make possible the assembly in the distant future of motion pictures which will depict and compare all phases of each major subdivision of the sciences.

While the conception of the truly scientific film is not yet complete, and the acceptance of motion picture film as a major means of recording facts permanently is not yet widely accepted, it is hoped that those who truly wish to serve science, even though they do not believe that motion pictures will achieve a place of major importance in scientific institutions, will accord posterity the opportunity of deciding the matter by creating libraries or reservoirs of motion picture film.

LITERATURE CITED

KIMBERLY, A. E., AND SCRIBNER, B. W. *Summary Report of National Bureau of Standards Research on Preservation of Records.* Washington, March 16, 1937.

WEBER, C. G., AND HILL, J. R. *Care of Film Slides and Motion Picture Films in Libraries.* Washington, Research Paper, 17, November, 1936. (N.B. of S.)

WEBER, C. G., AND HILL, J. R. *Evaluation of Motion Picture Film for Permanent Records.* Washington, Nat. Bureau of Standards, July 16, 1937.

Journal of the Society of Motion Picture Engineers., 27, December, 1936.

Regulations of the National Board of Fire Underwriters for the Storage and Handling of Photographic and X-Ray Nitrocellulose Films. New York, Nat. Board of Fire Underwriters, July 15, 1930.

Regulations of the National Board of Fire Underwriters for the Nitrocellulose Motion Picture Film. New York, Nat. Board of Fire Underwriters. August 15, 1931.

"*Vaporate Film Treatment to Protect Your Films.*" Filmo Topics, p. 11, Summer, 1937.

THE CHRISTIAN TODAS *

M. B. EMENEAU

Honorary Fellow in Linguistics, Yale University

(Communicated by Franklin Edgerton)

Abstract

A description of the Christian Todas in their relations with the flourishing culture of the tribe. They retain in the economic sphere a dependence on the tribal pastures, and show a tendency to retain the old type of economic partnership with individual Kotas. The converts retain their Toda names, even when these belong to the tribal religious sphere; in naming children born within the Christian community there is a tendency not to give Toda names. Cross-cousin marriage is followed when possible; disputes on matrimonial affairs follow tribal patterns. The tribal dance and song-technique are retained, and one Christian lyric of great beauty was recorded. The tribe treats the Christians as outsiders in general, but several links remain, notably in the meeting of obligations by both sides when marriages take place. There are attempts by the tribe to force the Christians to meet obligations on the occasion of tribal funerals; when they are met, it is generally due to compulsion. An attempt is made to evaluate the whole situation in terms of the attitudes of tribal members.

During a linguistic field-trip among the Todas of the Nilgiris in South India, I came into contact also with the section of the tribe who have become Christians. In fact, the two interpreters whom I used in my investigations into the Toda language were Christian Todas, since it is only among these that any fluent knowledge of English is so far to be found. A number of interesting episodes of Christian Toda life came to light during the period of nearly three years in which I had intermittent contact with the two communities. It seems useful to collect these and to attempt an evaluation of the elements of the tribal culture which have survived among the Christians and of the relations between the tribal Todas and the Christians. The attempt is worthwhile since studies of Christianized communities which exist by the side of the parent communities with the original culture in full working-order are rare.

The only mission which has worked among the Todas in recent times and made converts is the Church of England Zenana Mission Society. The work was begun in 1890 by Miss Catherine F. Ling [1] and seems to

* This paper represents a small part of the work which I was enabled to do in South India with the aid of the American Council of Learned Societies in 1935–6 and of the American Philosophical Society (Penrose Fund) in 1936–8.

[1] Two small books by this lady describe her work: *Dawn in Toda Land* (3rd ed., Madras, S. P. C. K. Press, 1919) and *Sunrise on the Nilgiris* (London, Zenith Press, ?1933). There is little of anthropological interest in these, which were of course written as descriptions of missionary work; the resistance of the tribe however is well brought out and some of the accounts of converts' histories amount almost to life-histories such as anthropologists gather.

have had as its immediate incentive an incident in which a Toda man who was ill was given medicine by a Tamilian Christian and told to pray at the Tamil Church in Ootacamund. He seems to have interpreted this as a vow to a god, such vows being freely undertaken by the Todas,[2] and on recovery he went to the church to do whatever rites were customary. The missionaries regarded this as a direct call to undertake mission work with the tribe and Miss Ling was appointed to be in charge. In 1904 the first convert was baptized, [kiške*r*] (59, Kishkar),[3] his own name being used as a baptismal name along with Paramajyoti, a Sanskritic name of the Tamil and Tamil Christian type. This man was appointed a catechist to preach to the Todas and still acts as such. In 1907 the next converts, five young women, were baptized. Preaching and teaching were carried on from the beginning in the face of recurrent opposition from the tribe, although much of the teaching was only indirectly religious. When converts were made, economic difficulties arose since these persons were cut off from the tribal life, both by the tribe and also, though reluctantly, by the missionaries, whose aspiration had been to convert the tribe as a whole and retain the old economic structure while doing away with the religious and the more un-Christian social features of the tribal culture (such, *e.g.*, as polyandry, the low status of women, etc.) and adding agriculture to the pastoral regime. The difficulty was met by getting permission to use some of the abandoned Toda village and dairy sites and pasturages; on two of these the converts were settled and agriculture was joined to the care of cattle in providing a living for them. At both places a church was built, the more important having a turret and bell provided by the converts themselves, which Miss Ling, probably correctly, regards as an equivalent in the life of the Toda Christians to the sacred bells of the tribal dairies. Houses for the people, schools, and medical dispensaries complete the material side of the mission colonies. The number of converts has varied. Many Todas at various periods have undergone religious instruction preliminary to baptism without being able to take the final step. At present the number of persons living at the two colonies is

[2] See my paper *Toda Culture Thirty-five Years after* (Annals Bhandarkar Oriental Research Institute **19**. 101–121), 117.

[3] When a Toda is mentioned, he is placed with reference to the genealogical tables in W. H. R. Rivers, *The Todas* (London, Macmillan and Co., 1906), by the number of the table in which his name occurs and the spelling there given. The writing of Toda words in this paper is strictly phonemic. The phonetic characters used are in general those of the IPA. Small capitals represent retroflex consonants, except that [ş] is used for the sibilant. Italics represent alveolars, [*r*] being an alveolar trill. [r] is a post-dental one-flap tremulant. [š] represents the abnormal sibilant (almost like English *sh*). [c] and [ʒ] are affricates of *ts* and *dz* type respectively. [ö] is a mid, mixed, rounded vowel, [ɯ] high, back, unrounded. When two vowel characters are juxtaposed, they together represent a diphthongal phoneme; [aːɯiː] in one name is the diphthong [aːɯ] plus the vowel [iː].

about 50, including some wives who were Tamilian by origin, and all the children.[4] In 1932 Miss Ling returned to England after fifty years of missionary life. The Toda Christians who had turned to her for settlement of all their difficulties now found themselves without strong and sympathetic guidance, and dissatisfaction with those who succeeded her and quarrels among themselves, as well as difficulty in meeting their financial obligations to the mission and to Miss Ling herself, who had aided them with large loans from her own means, produced in the community what may only be regarded as disintegration. In 1938 Miss Ling felt compelled to return to India and her influence had already begun to heal the breaches before I lost contact with the community.

A few of the men have taken to other economic pursuits than agriculture. One is a chauffeur, one a policeman, another was apprenticed to a tailor. [kuaʀoːn] (34, Kwodron) showed promise in school and later studied at the Madras Christian College in Madras, where he learned English with remarkable correctness and fluency (he was my best interpreter). He was employed as a clerk in various government offices in Ootacamund but got into financial difficulties of a somewhat disreputable sort, and finally to extricate himself became a convert to the Seventh Day Adventist Mission. [kaɴfɯsodj] (14, Kunpuradi) has been a chauffeur but is now an agriculturalist. His English is rather defective, but he made a remarkably good interpreter and informant in investigations into ethnological matters, since his interest in tribal affairs is intense, and in the song-technique of the Todas, with which he is thoroughly familiar, being himself the quasi-official composer of the Toda Christians. Most of the members of the community work industriously at their agricultural life and, since the Nilgiris specialize in growing potatoes for the markets in the plains and as far north as Bombay, economic life is fairly secure and permits a standard which is in many ways higher than that of the tribe and than that of many other Christian communities of South India. The women continue to do Toda embroidery for sale through the mission.[5]

Economically, the Christian community has had to separate itself from the tribe. One link however remains. The pastures of the Christians become poor seasonally and at such times cattle which are not giving milk are often pastured with the tribe's herds in the dry-weather pastures. The man to whom such cattle are entrusted is recompensed with a small fee. Frequently the same course is adopted with young

[4] This figure is only a guess and probably a slight overestimate. The census figure of 597 for the Todas in the year 1931 included 27 Christian Todas, according to the enumerators, although the mission claimed 37. The Christian community has undoubtedly grown since then by the addition of converts and of births.

[5] See my paper *Toda Garments and Embroidery*, JAOS 57. 277–289.

animals which have not yet reached the milch stage and which otherwise would unprofitably use up the community's scanty pasturage. It seems also that the Christians still retain some of their original feeling that wealth is to be measured by possession of cattle. They tend to acquire more and more head of cattle and because of the pressure on their own pasture area the superfluous beasts are kept with the tribe almost permanently or in alternation with unproductive animals. In these cases the man who provides pasturage gets the produce, milk and the resulting products. In all cases when cattle are pastured on tribal lands, it seems that it is the buffalo that is so treated; cows are kept at the Christian colonies, since the tribal Todas do not own cows.

The Todas live in economic symbiosis with the other aboriginal tribes of the Nilgiris, viz. the Kotas and the Kurumbas, and with the Badagas, who settled in the Nilgiris perhaps five centuries ago. The mechanism of the economic interchanges has not yet been described in detail,[6] but an outline of the Toda relations with the Kotas will be given here. The Kotas are blacksmiths and potters, and all the Toda needs in these departments are supplied by them, as well as music for the Toda funerals. In return the Kotas receive the carcasses of all Toda buffaloes that die or that are slaughtered at funerals, as well as certain other minor perquisites, such as cane for house-building, bamboo churning-sticks, and a certain amount of clarified butter. The interchanges are made between individuals, who are in a permanent and usually hereditary relationship to each other, referred to by the Todas as [keʟ], by the Kotas as [keʟ]. The relationship includes also the right of the Kota to obtain by purchase from his Toda opposite number such buffaloes and buffalo calves as he may need for his own economic purposes. In most cases one Toda has several Kotas who supply him, the relative numbers of the tribes and the Kotas' similar relationships with the Badagas making such an arrangement necessary. This economic relationship has broken down formally so far as the Christian Todas are concerned. But in spite of this, there is a tendency among the Christian Todas to retain the relationship with their hereditary Kota partners. They obtain their iron tools and pots from these Kotas and recompense them on a more strictly commercial basis than the original one. *E.g.* [kaɴfɯsodj] retains such relationships with the Kota who was his father's partner and has in addition formed a new relationship with another Kota. He at least still uses the term [keʟ] in referring to his relationship with these two men, though my Kota linguistic informant was insistent that the Kota

[6] Reference may be made to Rivers, chap. 27. Dr. D. G. Mandelbaum has made more extensive notes on the subject and it is hoped that his account will soon appear.

term [keʟ] should not be used of such relationships with the Christian Todas.

In religion there has of course been a complete break. None of the tribal religious practices could have been fitted into the Christian pattern and there has been no attempt to do so, but rather a revulsion on the part of the missionaries from anything that smacked of the old religion (note however Miss Ling's acceptance of the attitude towards the church-bell mentioned above). This has gone so far that the missionaries who succeeded Miss Ling frowned on dancing. In the tribe the dancing is a part of every social occasion. It is performed at religious ceremonies, but also at funerals, ceremonies when the ears are pierced, weddings, and almost all other occasions when a number of the people gather. Even the children dance in play in the style of their elders (girls have been seen dancing as well as boys, though Toda women never dance). Miss Ling, realizing that the dance is not religious, looked with an indulgent eye on the dancing of the converts and in fact rather encouraged it than otherwise, with its accompanying extempore songs, at weddings, christenings and other joyous occasions. The missionaries in charge during the interregnum when Miss Ling was in England, with lack of insight and sympathy, forbad dancing. The reason given was that such a typically tribal activity would produce a hankering for other tribal customs and consequent backsliding from Christianity; possibly also a Puritanical attitude to dancing in general played its unacknowledged part. The effect that I observed however was that when the missionary guests retired from the scene of festivity, wedding or the like, dancing took place, and there resulted a regrettably un-Christian fostering of deceit and dislike of the missionaries in charge. It is to be observed that the Christian festivities are almost always attended by tribal Todas of the neighborhood and also by others from a greater distance who are related to the principals in Christian weddings and christenings. They are participants in the feasting and the males among them join with alacrity in the dancing.

The song-words that are composed during the dancing, as well as all other songs in the Toda language that are composed by the Christians, are of the tribal type and use the normal Toda technique of word and tune.[7] As most Toda songs are extempore and topical, so are the songs made by the Christians. There is an attempt to enlarge the verbal technique to embrace specifically Christian concepts and considering the rigidity of the conventions there are signs that the attempt will be suc-

[7] See my paper *The Songs of the Todas*, Proceedings American Philosophical Society **77**. 543–560.

cessful. Miss Ling had expressed the hope that the converts might some day break into Christian song in their own tongue and especially that a Benedicite might be composed. It is obviously too much to hope for that there should be a completely new style of song evolved; all that one could reasonably expect would be either that on the basis of Tamilian music or, less likely, European music songs would be composed with Toda words, or that with practice the conventions of Toda singing could be extended to cover the Christian spirit and subject-matter. With a view to helping on the latter evolution (perhaps ethnologically unethical, but certainly justified if it would contribute to filling an evident gap in the life of the Christian Todas, and also interesting as an experiment to see how far the technique could be made by a clever composer to cover a previously untreated subject), I suggested to [kaɴfɯsodj], my interpreter and a man versed in the song-technique, that he should attempt a Benedicite. He without further prompting attempted composition in both the modes suggested above. His fairly literal translation, based on his Tamil Prayer-book and to be sung to the tune used in church, was hardly a success. But since he had sharpened his wits by this effort, there was a much more satisfactory result. At the end of a night when he had been sitting in his potato-field to prevent raids on his crops by wild pig and deer, as dawn broke he burst forth in a song. This he later dictated to me and it proved to be a remarkable blending of the Christian subject-matter and the imagery suggested by the Toda poetical technique. The song, when read in translation to westerners whose judgment in such matters I value highly, has been adjudged a highly poetical lyric. Certainly for poetic value it far surpasses most of the 300 Toda songs that I have recorded. Whether efforts on such a high level can be produced in greater number remains for some future investigator to discover.

I have shown elsewhere [8] that the personal names of Toda males are in almost all cases derived from the religion of the dairy-complex and more specifically from the complexes of the patrilineal sib in the case of each man, while the names of Toda females are more miscellaneous but in many cases show an origin in the song-language. In all cases the Toda Christians who were proselytized from the tribe have retained their Toda names, usually as baptismal names, in spite of the definitely religious connotations in the case of male names. In addition they have all received Christian names, either English or Sanskritic-Tamil. When we come to the children born to the Christian community the case is different. There the Christian names are in the foreground of interest and are the only ones given at baptism. In most cases it seems that no

[8] *Personal Names of the Todas*, American Anthropologist, N.S., 40. 205–223.

Toda names were given at all. *E.g.* [kaɴfɯsodj] gave no Toda names to his children, though his interest in tribal affairs is intense and his connexion with the tribe is fairly intimate. I cannot explain his failure to give Toda names, though a contributing factor is probably to be seen in the fact that his wife is not Toda by origin. In the few cases where I have recorded Toda names for the children of Christian Todas, tribal patterns were followed for the females. *E.g.* [tuaːsmonguɯːʀ] (59, Tormungudr) named one of his daughters [sotjabuːf] "spiritual power-flower," [9] which has as its first member [sotj], occurring in the song-language paired with [tuːʟ] in references to the power of the gods or the dairy-complexes. There is probably in this name however a transfer to the Christian sphere of thought. This man's other daughter was also given a Toda name, but she is so generally known, even to the tribal Todas, as Mary that my informants were unable to remember the Toda name. Two girls recently born were named [piʟjiːʒfuːf] "silver-flower" and [tyʟjiːʒfuːf] (the first member [tyʟj] is not clear).[10] One name of a male child belongs to the tribal religious sphere; [kuaʀoːn] has a son called [asojxuɯɪᴛɴ] from [asojɯr], the buffaloes of the [tiː]-dairy of this man's original patrilineal sib.[11] Another son of his is called [sinxuɯːʀ] "gold-horn," which has no sib-connexions; the formation resembles that of women's names in its use of [sin] "gold," but the second member [kuɯːʀ] "horn" is an element found not in women's names, but in male names with sib-connexions.[12] Similarly with [pumxuɯːʀ] "fruit-horn,"[13] the name of a son of [iːṣxuɯːʀ] (21, Idrshkwȍdr).

Tribal social customs have had in general to be abandoned. Funerals and weddings, for example, are conducted in Christian fashion, and the tribal manner of legitimizing children by giving a toy bow and arrow to the pregnant mother can no longer be used. A survival of tribal custom may be seen however in a preference for cross-cousin marriage (which is not prohibited by Church of England canon law or by English civil law). This is the preferential form within the tribe but operates with many exceptions. Consequently, marriage among the Christians can take place between those who are not cross-cousins without departure from tribal patterns. The small number of the Christians makes non-cross-cousin marriage obligatory in most cases. One case, however, is known to me in which cross-cousins were available, and here the failure

[9] *Ib.* §§ 26–8, for the formation.

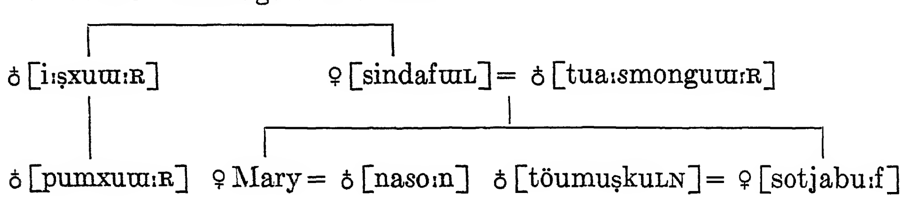
[10] *Ib.* § 24 for -[iːʒ] and § 27 for a name with this suffix added to its first member and followed by a second member.

[11] *Ib.* § 9.

[12] *Ib.* § 8.

[13] *Ib.* § 28, for the element [pum].

of the marriage to take place was the occasion of quarrelling and attempts to penalize the person responsible for the failure. [iːṣxuɯːʀ] (21, Idrshkwôdr) has one son [pumxuɯːʀ]. His sister [sindafɯʟ] (Sindavul) married [tuaːsmonguɯːʀ] (59, Tormungudr)[14] and they had two daughters, Mary and [sotjabuːf]. [iːṣxuɯːʀ] wanted Mary to marry his son, her cross-cousin, but since the boy was not old enough, Mary was married to [nasoːn] (son of [teːpax] 20, Teipakh). (In the tribal system [nasoːn] and [sindafɯʟ] would use brother-sister terms to one another, since they were classified thus in the same patrilineal sib, but Mary could have been married to [nasoːn] in spite of that, since she would have been of another patrilineal sib and not related to him through a matrilineal sib; no violation was done by this marriage either to the tribal system or to the Christian system.) On the occasion of this marriage it seems that the girl's father paid a fine of one buffalo (or cow, I am uncertain which) to [iːṣxuɯːʀ], a clear carry-over from tribal custom, and promised that when his other daughter grew up she should be married to [pumxuɯːʀ]. When she came of age however, [pumxuɯːʀ], who is a ne'er-do-well, was unable to support a wife. Her father consequently went back on his promise and made arrangements for her to marry [töumuṣkuʟɴ]. [iːṣxuɯːʀ] tried to have the publishing of the banns stopped and even went in his appeals as far as the Bishop of Madras. The missionaries, who have no objections to cross-cousin marriage, helped [pumxuɯːʀ] by giving him cattle with which he might start his household, but he sold them and squandered the money. Under the circumstances nothing was left but to let the marriage of [töumuṣkuʟɴ] and [sotjabuːf] take place. The Toda Christians then held a formal meeting, assisted by one of the tribal Todas, [peːθoːʟ] (7, Peithol), and he handed down a judgment with the consent of the majority that the girl's father should pay a fine of three cows to [iːṣxuɯːʀ]. This was a purely tribal solution and when the missionaries heard of it, they objected vigorously. The community decided however that the fine must be paid and a story was invented that [tuaːsmonguɯːʀ] owed two cows in the normal course of events to [iːṣxuɯːʀ]. These two were paid over and the third was smuggled along with them without the knowledge of the missionaries.



[14] A marriage not permissible within the tribe, since the parties were of different moieties.

This incident, without the missionary influence, is typical of tribal matrimonial disputes; in the tribe there would have been no objection to the cross-cousin marriage taking place as a formality to be succeeded when the parties grew up by a *de facto* non-cross-cousin marriage with a fine known as [ter] paid as it was in this incident. The formal marriage was barred by Christian custom, but the succeeding events follow the tribal pattern of [ter]. It is noteworthy also that a respected man of the tribe was called in to adjudicate; this seems to be usual with the Christian community even in cases where disputes are not so closely bound up with social customs.

We have seen then that in spite of the outward break with tribal culture some elements of this culture have been retained by the Christian community. The tribal pastures are still resorted to by the Christians when there is too much pressure on their own pasturage, and in some cases at least hereditary economic partnerships with Kotas are kept up and even new ones formed. Economic relationships with the Badagas were of a different nature, consisting mainly of a sort of rent paid by the Badagas to the Todas; these relations have disappeared completely with reference to the Christian Todas. So also, so far as I know, the relations with the Kurumbas, who figure in the economic symbiosis chiefly as sorcerers, one of whom is paid by each Toda, Kota, and Badaga village to keep all other Kurumba sorcerers at bay (a crude insurance system known elsewhere in India against both castes of thieves and castes who are reputed to be sorcerers). Converts retain their Toda names, even though in the case of men these have tribal religious significance. The children born within the community, however, in general have no Toda names; when such names are given those of females follow the tribal pattern with its frequent poetic associations and those of males are in a measure assimilated to the female pattern with at least one quasi-suffixal element retained from the male pattern. The matrimonial incident recorded in the last paragraphs shows that there is an almost complete carry-over of tribal patterns in this sphere. Finally, a most important element is the retention of the tribal dance and song-technique. In the latter we have a verbalization of thought and emotion expressed in the old familiar technique, with fairly successful attempts to adjust the rigid conventions to the new patterns, resulting in one case in a highly poetical expression of the Christian attitude to God in his role of creator and being to be worshipped by all creatures.

The Christian Todas can hardly be said to have reached a perfectly integrated state. It is interesting to speculate on what their future condition will be, though very unprofitable. The adult members of the

community are still in many respects Todas. The new generation however is growing up without intimate knowledge of things Toda and in most cases without knowledge of the Toda language, since there is an admixture of Tamilians in the community and all formal instruction is carried on in the Tamil language. New converts may and probably will be made who will introduce fresh Toda influences. But it is to be expected that the community will in time become so Tamilized that those born into it will hardly differ at all from other Tamil Christians and fresh converts will find few Toda traditions there to strengthen.

It still remains to discuss the attitude of the tribe towards the Christian community. Theoretically since the Christians have "broken caste," have given up the tribal religious and social rites and have adopted some habits which are repugnant to the tribe, *e.g.*, meat-eating,[15] there should be an absolute break in the relations between the two communities. The chief marks of outcasting in most Hindu communities are refusal of the caste to intermarry with the outcastes and also to interdine. The latter is based on the conception of pollution; a person who is not of the caste and who does not observe the customs of the caste, especially in such intimate matters as preparation and consumption of food, will produce equal uncleanness in the caste-members, necessitating elaborate purifications. The Todas are not in a strict sense a caste, though conceptions of purity and pollution are prominent in their culture. They will not, for example, go into a Kota village since the Kotas are unclean by occupation (blacksmithing and pottery) and by eating habits (meat and even carrion are regularly eaten); all transactions with Kotas at a Kota village must be carried on at the green outside the Kota village. Nor is a Kota allowed to enter a Toda village. Interdining between Todas and Kotas is non-existent. The Badagas however do not suffer from the restrictions that the Kotas do. Badagas may enter Toda villages and those in certain formalized relations with the Todas are invited to Toda ceremonies and eat with them at the feasts that follow. The European is on about the same footing as the Badaga, though he eats meat. I have been invited on several occasions to take part in the feast following a Toda ceremony and my sitting in the row of Toda feasters did not inconvenience them as that of a Kota would. The Toda Christian also is invited to eat at such feasts and does eat. He is not however allowed into a Toda dairy, any more than any other profane person is, nor is he allowed into a Toda house, in this respect being like the European. The European may enter the

[15] See *Toda Culture Thirty-five Years after*, p. 119, for the somewhat equivocal attitude of the tribe to this habit.

house if he takes off his shoes and if the cooking utensils have been removed from the house; this seems to be a mark of courtesy to the ruling caste and no Toda will trouble himself to remove the cooking utensils to oblige a Toda Christian. This, I think, is the essence of the tribal attitude. The Toda Christian is an outsider, with the complication that he was once a member of the tribe and still has all the links of blood-relationship and affinal relationship with the tribe. There is resentment because he has left the tribe, and on occasion the diviners *ex officio* and even other members of the tribe may charge the misfortunes of the tribe to the fact that there have been converts to Christianity and may commence open hostilities (wordy quarrelling and sometimes blows) if a Christian Toda happens to be present, especially if he has brought a European to see ceremonies. At the same time the pull of the original relationship survives in a measure and is clearly evident when a respected member of the tribe sits in the Toda Christian meeting and gives the benefit of his advice and when those members of the tribe related to the principals at Christian weddings and christenings attend these ceremonies.

The overt actions of the tribal Todas can be illustrated in more detail. The other side of the interdining relation works as follows. The tribal Toda when he comes to a Christian Toda house or ceremony will eat only if the food has been cooked by a man, either tribal in origin or born in the Christian community; this is a little obscure since food is ordinarily cooked in a tribal Toda house by the women and the question of menstruation pollution does not enter to complicate the picture.[16] It may be due to the presence of women of Tamil origin in the community or perhaps to the fact that women converts are not readmitted to the tribe even if they wish to leave the Christian community. Furthermore, the cooking must be done in a metal vessel and not in an earthen one; according to general Hindu views pollution can be removed from the former but not from the latter.

Readmission to the tribe of a Toda Christian male who was born in the tribe is always possible and has happened often enough. For example, [kaɴfɯsodj] went back to the tribe on one occasion in order to marry a woman with whom he was in love. He was unable through illness to stand the physical life of the tribe, however, and became a Christian again. It appears that the chief ceremony in readmission to the tribe was provision for him of a new perineal cloth [kuɯːɴ], the fundamental garment of the Toda man.[17] A Toda woman, once she

[16] See my paper *Toda Menstruation Practices*, which will appear in the Dr. F. W. Thomas Festschrift Volume to be published by the *New Indian Antiquary*.

[17] Cf. the regulation that a man who has lost his perineal cloth is barred from going to a dairy until after sunset of that day.

has been baptized, will not be readmitted to the tribe. Children born in the Christian community are not admitted to the tribe; the reason given is that they are not legitimate, since no toy bow and arrow has been given for them, and so they cannot be fitted into the tribal social structure. [alc] (3, Ultz), his wife [sirmɯl] (30, Sirmul) and their children attempted to return to the tribe; only the man was received.

Blood and affinal relationships within the tribe involve each member of the tribe in a network of rights and obligations. At a funeral, for example, certain people related to the dead person must provide buffaloes to be slaughtered, and every man in the tribe has obligations of this sort which he will sooner or later have to meet, while on the other hand when he dies it will be obligatory on others to provide buffaloes for his funeral. It is felt by the tribe that the Christian Todas still have obligations of this sort towards their relatives in the tribe. Though the missionaries object to the obligations being met for funerals, the slaughter being based on a non-Christian attitude towards a future life, it has happened on several occasions that the Christians have met their obligations, either willingly or unwillingly. [kerxuɯːʀ] (63, Kergudr) provided a buffalo for the funeral of his father [kanɯs] (Kaners); I do not know whether this was voluntary on his part or not. In the case of [kaɴfɯsodj] however my notes are more complete. When he was a recent convert, on two occasions he provided buffaloes for classificatory brothers, [pɯnog] (14, Punog) and [kiškeːn] (14, Küshken). He says that he thought it best to follow the tribal rules and to yield to the insistence of the tribal council. Recently, however, [nertnɯs] (24, Nertiners) died, who had married [paːɯiːʒfiʟj] (14, Poidjveli), the classificatory sister of [kaɴfɯsodj]. Her brother [noːnmuʀj] (Nonmudri) and [kaɴfɯsodj] were the only surviving affines of [nertnɯs] who could be called on to provide a buffalo for the funeral; according to the present regulations they should have provided one buffalo between them. [noːnmuʀj] had no buffaloes. One of the animals belonging to [kaɴfɯsodj] was at a tribal pasture under the care of [idjoːn] (20, son of [teːpax]). This was seized and slaughtered at the funeral in spite of protests by [kaɴfɯsodj]. It was finally arranged by the tribal council that [noːnmuʀj] should pay to [kaɴfɯsodj] half the value of the buffalo.

Other obligations are found on the occasion of a woman's first marriage. Her husband gives to her father his best buffalo or two buffaloes, and five rupees; these are the [kökuɯlj] "bribe." He also makes a feast for all the Todas. The father then gives the husband [aʀpɯn] "dowry," *i.e.* a number of buffaloes, at least twice as many as the "bribe," plus one, *i.e.* three if the "bribe" was one, five if it was two.

He also gives a feast. If the father is not alive, the duty devolves on his brothers, real or classificatory, in his immediate family, then on the woman's brothers, real or classificatory, in the immediate family, then on the generation of her brothers' sons. This obligation, both in regard to "bribe" and to "dowry," is still maintained in most cases between the two communities. The missionaries in this case approve, at least, Miss Ling approves. About ten years ago the Christian [kišker] gave one buffalo to [keːnoʒ] (21, Keinodz), who is not a Christian, as "bribe" for the latter's classificatory sister [sodamut] (Sadamut), who was already a Christian; [keːnoʒ] then gave three buffaloes as "dowry." Previously the non-Christian [idjoːn] had given "dowry" for his Christian sister [lamdiːfj] (20, Numdeivi) when she was married to [kišker], the "bribe" having been given by the latter. When [kišker]'s father's sister [tuarfɯʟemj] (59, Turvalami) was married to [piθjoːnaʟf] (64, Puthion), both the latter being non-Christians, a "bribe" was given to [kišker] by the latter and [kišker] gave three buffaloes as "dowry." [kuaʀoːn], who got his wife from the Todas by carrying her off (the [ter] pattern), did not give a "bribe" or get "dowry," though it is clear that the tribal patterns demanded it.

We find then that the tribe, though it has some resentment against the Christian converts from its ranks and treats them in some respects as outcastes, yet retains with them some rather important relationships. In marriage affairs the Christian Todas tend to be treated, without compulsion on either side, as if they were still members of the tribe. In funeral matters the tribe tries to demand that the Christians should fulfil their obligations, though the Christians are on the whole unwilling to meet them. If we look at the Badagas, we find that Christian converts from their ranks are outcastes and that the original community has no relations at all with them. No Kota has ever become a Christian. When we go to the more classical Hindu areas, we find from missionary accounts at least (few other accounts have ever been given), that in many cases the Christian converts are entirely outcaste and that no relations are maintained with the castes of origin other than that of extreme hostility on the part of the caste (the picture is perhaps not so simple as this, but accounts are defective). Can any explanation be given for the peculiar features in the Toda case? It is possibly to be found in the general situation. The areas in which the tribe and the Christians live is comparatively small, the two communities see much of one another necessarily, the Christian community still retains cattle as a necessary part of its economic life and so has a community of interest with the tribe and moreover has had to depend on the tribe in part for

its pasturage, to the economic advantage of the latter. Moreover, the tribe has come to depend on the mission for the sale of its women's needlework. Perhaps another factor is to be seen in the small size of the tribe. Every convert is related somehow to practically every Toda in the tribe and still uses the appropriate terms of address. The Badagas show a contrast here, for in a community of over 40,000 the great majority of Badagas that a convert will meet will be strangers and unrelated to him. It would be unwise to push any argument based on this feeling of relationship between the two Toda communities, and yet one suspects that the matter has its weight. Finally, if one must, it is possible to fall back on characterizations of the Todas as "a friendly people"[18] or as a tolerant people. Their hauteur and self-sufficient bearing has often been remarked, and as a complementary side of their character there is a lack of fanaticism in them. They know that they are a peculiar people (in anthropological terms, aberrant) and recognize that others, even if they are Todas, must be allowed their peculiarities. Hence their rather great degree of tolerance for the Christian converts and their retention of relations with them where it is possible.

[18] So my Kota informant quoted in *Toda Culture Thirty-five Years after*, p. 105.

SYMPOSIUM

PROGRESS IN ASTROPHYSICS

JOINT MEETING OF

THE AMERICAN PHILOSOPHICAL SOCIETY

AND

THE FRANKLIN INSTITUTE OF THE STATE OF PENNSYLVANIA

FEBRUARY 17, 1939

PROBLEMS OF THE SOLAR ATMOSPHERE

DONALD H. MENZEL

Harvard College Observatory

(Read February 17, 1939, in Symposium on Progress in Astrophysics)

THE sun is the recognized "proving ground" for the testing of astrophysical theories. In the absence of information to the contrary, we are forced to regard the sun as a representative star. The peculiarities of structure exhibited by the sun must be regarded as the possible feature of any star, whether the star is similar in type to the sun or not. The solar "proving ground" has been rather severe, and few theories have emerged unscathed from the rigorous tests of observation.

In certain respects the high mortality of astrophysical theories is not without compensation. For example, almost a dozen different hypotheses have been developed in an attempt to explain one single phenomenon—levitation of the solar chromosphere. If it were not for the rigorous restrictions of the detailed observational data, we should now be faced with the difficulty of trying to decide between the various theories, any one of which might serve as a qualitative explanation of solar phenomena. Instead, no theory as yet proposed seems adequate to fit all the observational data. I consider the latter alternative, no theory as opposed to a dozen, to be distinctly healthier for progress in astrophysics. It is often regarded as beneficial for the young to scrape for a living. Astrophysics, a young science, should not find its path too smooth.

Problems of the solar atmosphere may be divided rather arbitrarily into two different groups, (1) the interpretation of solar spectra, and (2) motions and fields of force in the solar atmosphere. There are, of course, still other aspects of the problem—and even these two overlap. After a very brief summary of the present status of (1) I plan to devote the major portion of the present paper to (2), returning to consider (1) only in so far as problems of spectroscopic excitation are allied to those of atmospheric motion.

During the past year or two a number of investigations relative to the Fraunhofer spectrum have appeared. Quantitative measures of the equivalent widths of selected solar absorption lines have been made, chiefly by Minnaert and by Allen. The interpretation of these measures

by means of the theory of the "curve of growth" has been progressing satisfactorily. Through the medium of the curve of growth, the observed line intensities may be translated into numbers proportional to the numbers of atoms in the lower of the atomic levels from which the absorption line arises.

Comparison of relative populations of levels that differ in excitation potential allow determination of the effective "excitation temperature" of the atmosphere. The calculations of effective temperature are based on the well-known Boltzmann equation of statistical mechanics. Strictly speaking, we may apply this equation only when the gaseous assembly is in thermodynamic equilibrium, *i.e.* when a single parameter, the temperature, serves to define the degree of excitation. The very fact that the solar spectrum shows absorption lines—even the fact that the sun possesses a spectrum at all—shows that thermodynamic equilibrium does not and cannot exist in the sun. The ideal condition of thermodynamic equilibrium cannot prevail in a stellar atmosphere through which energy is flowing. It can exist only in a perfectly insulated chamber.

We may, nevertheless, employ the Boltzmann equation symbolically, and use the "excitation temperature" as a convenient parameter to describe the physical condition of the assembly. We recognize the possibility of having to use a different parameter for each atomic energy level, possibly one for each spectral line. Fortunately, nature seems to avoid making it too complicated. A number of independent analyses are now available, covering the spectra of many different substances, including those of several molecules. A wide range of excitation potential has been investigated. Although departures from the conditions of thermodynamic equilibrium undoubtedly exist, the results are fairly consistent in indicating that the departures are small—at least for the metallic atoms. A temperature of about 4,500° has been determined for the reversing layer.

The lowness of the temperature led to considerable comment at first. The difference between this figure and the value of 6,000°, ordinarily accepted as the solar temperature, is rather startling. If we are to interpret the new figure as the true temperature, it would seem to be more nearly representative of the boundary than of the low-lying photospheric layers.

Departure of the solar atmosphere from the conditions of thermodynamic equilibrium is a topic that has become increasingly difficult to discuss. As we consider the entire solar atmosphere, we see ample observational evidence, proving conclusively that a single temperature

cannot be employed to represent the physical state of the gaseous assembly. The outstanding and most frequently cited example is the occurrence of the lines of helium and especially of ionized helium in the spectrum of the chromosphere. The order of magnitude of the helium discrepancy greatly exceeds that for the other elements. Nevertheless, there appear to exist similar discrepancies, both in the Fraunhofer and chromospheric spectra, for the ionized atoms of silicon and magnesium. Hydrogen, also, appears to be subject to the same anomaly.

There are two possible sources of departure from thermodynamic equilibrium. The first, which has been the one most widely discussed, is that resulting from the "dilution" of the radiation incident on the atoms of the solar atmosphere. The stream of radiation, according to this argument, is essentially unidirectional, over a hemisphere instead of over a sphere, as for thermodynamic equilibrium. Processes involving radiation, such as ionization, excitation, etc., will proceed, therefore, at one-half the normal thermodynamic rate. The converse processes of electron capture, spontaneous relapse of excited atoms to the ground level, etc., will proceed at the usual rate. In consequence we should expect to have a concentration of atoms in the ground level amounting approximately to a factor of two.

This simple view of the departure of an atmosphere from thermodynamic equilibrium failed to meet the tests of both observation and theory. It is in the wrong direction, of course, to explain the phenomenon of superexcitation. Although the effect is in the right direction to explain away part of the 4,500° anomaly of the reversing layer, the observations show no definite indication of an excess in the lower atomic levels. Failure of the effect to show up may be ascribed, in part, to the operation of collisional excitation. But the primary cause, I believe, is that the dilution factor of one-half does not exist. In the actual process of forming an absorption line, an appreciable two-way stream of radiation is set up. The unidirectional stream, which gives rise to the theoretical dilution factor of one-half, occurs only in wave-lengths that are not being absorbed. The close conformance of the excitation of a single atom to the conditions of thermodynamic equilibrium may be due in part to the fact that the total radiation flux contributing appreciably to an absorption line, is only a small part of the total radiation flowing through the given atmospheric level.

Study of the conditions of excitation in the gaseous nebulæ shows that the *quality* of radiation, *i.e.*, the spectral distribution, is fully as important as dilution in determining the conditions of excitation. And most important in this connection is the intensity of radiation within

the ultimate lines of the atom in question. The dilution does have a pronounced effect in increasing the relative populations of atoms in the ground and metastable levels, but the population of atoms in the excited levels will depart but little from the conditions of thermodynamic equilibrium. We should expect a similar effect in the atmospheres of stars. Milne and others have given extended ionization equations, averaged for the different pressures existing in an isothermal atmosphere. It would seem that the effects of temperature gradient, as discussed by Max Krook, might be even more important. Further, the distortion of the energy curve of a star, *i.e.*, its departure from the black-body law, must also exert a major influence upon the conditions of ionization and excitation.

The chemical composition of the sun is of great importance. We are faced with a double problem. First of all, we must meet the challenge of deducing from the intensities of the Fraunhofer lines the absolute as well as the relative numbers of atoms of each element above the photosphere. To carry out these calculations in detail we require advance information concerning the absolute absorbing power of an atom. Although this quantity is known for hydrogen and helium, and can be closely estimated for other atoms, considerable study remains before we can obtain entirely reliable estimates from quantum mechanics, relative to the absorbing power of complex atoms. Nevertheless, the order of magnitude of the absorbing power can be estimated.

The second part of the problem is, given the absolute number of atoms of each variety above the photosphere, calculate the location of the photosphere from a knowledge of the general opacity of the atmospheric atoms. Various sources of opacity for the continuum have been investigated. Pekeris and Menzel had calculated the effect for bound-free and free-free transitions, both in hydrogen and for complex atoms. Their formulæ, however, when checked with the observed number of atoms of all varieties above the photosphere, fail to provide sufficient opacity. Wildt has suggested that the opacity may conceivably arise from absorption of radiation by negative hydrogen ions. Menzel, applying Jen's theory of the absorption coefficient of the negative ion, has shown that the discrepancy may be removed on the assumption of an abundance of hydrogen five thousand times greater than that of all the metals combined. Unfortunately, the problem of general opacity is related to questions of high excitation. It is impossible, at present, to give a definite answer. The relative abundances of oxygen, nitrogen, and helium are similarly affected. The abundance of helium, in particular, cannot at present be estimated. The calculated

abundance is markedly influenced by the adopted temperature. A relatively small uncertainty in the temperature makes all of the difference between having helium predominantly abundant, even over hydrogen, or having it a relatively rare substance. The indications seem to be that the solar atmosphere is largely hydrogen, with oxygen and helium somewhat less abundant.

The question of superexcitation leads us again to consider problems of the chromosphere, prominences, and the corona. The solar surface is far from uniform. Even as viewed in white light, the surface is seen to possess a fine granular structure projected against the dark background. Under high telescopic power the solar surface seems to be composed of granules, giving rise to the "network" discussed by Janssen and others. The network appears to be composed of myriads of tiny nuclei grouped together. Each nucleus, which appears as a tiny bright area, has an apparent diameter of less than a second of arc or about 800 km. Somewhat larger estimates have recently been given, but the most convincing observational data point to the smaller values.

From the fact that the granules are most conspicuous near the center of the solar disc, we conclude that they are low-level phenomena. In addition to the solar reticulation, we have the faculæ, bright areas often associated with sun-spot groups. These are best seen near the solar limb, which leads us to suppose that they are high-level phenomena.

The granulations and the faculæ, taken in combination with the even more prominent feature of the solar surface—sun-spots, lead us to believe that the sun's atmosphere is a complex phenomenon. We should scarcely expect to be able to explain its most prominent features by any theory that started with the assumption that the atmosphere is stratified in plain parallel layers. And yet this is the basic hypothesis of numerous theories of the problem of transfer of radiation through the solar atmosphere. At best, we should expect this assumption to hold over only relatively small regions.

Observation of the Fraunhofer spectrum indicates further complexities. As Evershed has shown, there appears to be a definite outflow of material from a sun-spot, evidenced by the Doppler effect. St. John has made a study of the similar effects for the solar atmosphere in general. St. John has found that there exists a correlation between the apparent Doppler radial velocity of a Fraunhofer line and its absolute intensity.

From measures of the Doppler displacement, St. John finds that the absorption lines of neutral ions, most of them of medium intensity, show a small displacement near the violet, which may be interpreted as

representing a slight outward velocity. Absorption lines of medium and high intensity show Doppler shifts in the reverse direction, indicative of a descent velocity that increases with the intensity of the line. The observed velocities range from 0.1 to about 0.7 km per sec. St. John also noted that the observed velocities show a correlation with the "heights" observed by Mitchell for the chromospheric extent of the various lines. The interpretation that has been given these results is that the velocities observed refer to the actual velocities existing in the various chromospheric levels. Evershed's sun-spot observations had suggested an outflow of gas. The analogy was drawn between the outflow from a sun-spot and that from a terrestrial volcano, as in the eruption of Katmai, where the ejected volcanic dust spread through the earth's entire envelope and took several years to settle out. The analogy, which tried to relate the slow fall of the volcanic ash to the descent of atoms over the solar disc, appears to be badly strained. A velocity of 0.1 km per sec must be regarded as a powerful vertical wind.

The correlation of the velocities with chromospheric heights has also probably been overemphasized. It seems reasonable to suppose that spectral lines observable to high levels in the flash spectrum might tend to have somewhat higher effective levels of absorption than the lines of less chromospheric extent. But there is no marked stratification within the chromosphere itself. There is a slight tendency for certain elements to possess intensity gradients higher or lower than the average. Rare-earth ions, for example, show a tendency to settle out, whereas the lines of helium show the slowest gradient of any atom, so that helium lines are relatively stronger at the highest level. Further observational data are needed to supplement the very excellent material obtained by St. John relative to motions within the solar atmosphere.

Although the motions within the reversing layer are still only partially understood, during the past few years much has been learned concerning the motions within the chromosphere and prominences. We owe this great advance largely to the development of the spectroheliograph at the McMath-Hulbert Observatory. The cinematographic record of the various types of prominences portrays more graphically than words or mathematical formulæ the nature of the problems associated with the motion of atmospheric gases. I am deeply indebted to Dr. Robert McMath for many interesting and enlightening discussions on this problem. These observations have had a most important part in forming my conclusions concerning the nature of forces operating in the solar atmosphere.

Observations of the flash spectrum and of solar prominences in general give us the information that some force or combination of forces is acting to support the atoms against the downward pull of gravitation. Many theories have been advanced concerning the nature of these forces, but our progress in interpreting the physical conditions in the chromosphere has been disappointingly slow. At our present stage of knowledge, the problem of chromospheric support is not dissimilar to the question of levitation of objects at a spirit seance. Both phenomena have been observed and most scientists agree that the agencies are not supernatural. How does the sun manage its trick of table-tipping? Its seances are held in broad daylight or in the partial gloom of a total eclipse.

Some years ago Milne investigated the role of radiation pressure in levitating atoms of ionized calcium. Difficulties entered, however, when quantitative application was made. Radiation pressure proved to be incapable of supporting more than a fraction of the total weight of the chromospheric atoms. The chief stumbling block in the way of developing a theory of chromospheric support based on radiation pressure was the levitation of atoms of hydrogen and helium, which can absorb only in the far ultra-violet, where the intensity of radiation from a black body at a temperature of 6,000° is negligible.

The observational data themselves, the general physical appearance of the chromosphere, with its spikes and prominences, suggested that the problem of chromospheric support was not a static phenomenon as postulated by Milne, but a dynamic one. Accordingly, McCrea advanced the hypothesis that the atoms reached high levels in the chromosphere because they had acquired high velocities at low levels. For the purpose of calculation only, it was assumed that these turbulent velocities were distributed according to Maxwell's law. There is, of course, no justification for this assumption, though the general effect of turbulence upon the chromospheric support is not invalidated because of this approximation. It may be objected that McCrea's theory does little to advance the problem since it makes no attempt to explain the origin of the turbulence. Such criticism is unfair; if the chromosphere can be entirely explained by turbulence, the difficulty is pushed one step farther back. The observational data indicate that the McCrea variety of turbulence does not exist in the chromosphere. Nevertheless, the random, heterogeneous motions of the atmospheric gases possess certain characteristics of turbulence.

The possibility that the spectral distribution of the solar radiation in the far ultra-violet may depart very markedly from that of a black

body at 6,000° lends some hope to the possibility of explaining chromospheric levitation on the basis of radiation pressure. It would seem reasonable to suppose that the observed inequalities of surface brightness exhibited in the solar reticulation might well be intensified in the extreme ultra-violet, and thus give rise to radiation-pressure effect.

Chandrasekhar, apparently of the same opinion, has re-investigated the problem. He dodges the fundamental difficulty by assuming, in advance, that the chromosphere, hydrogen and all, is supported by radiation pressure. He does not enter into a discussion of the question as to how this support is accomplished.

Several of the mathematical consequences of Chandrasekhar's theory are very simply derived. If an atom is subject to an upward acceleration r-times that of gravity, its velocity, v, and height, h, above the surface will be:

$$v = \sqrt{2rgh}. \tag{1}$$

Chandrasekhar assumes that $r = \frac{1}{3}$. Therefore, if an atom rises vertically to a height of 10,000 km, its velocity at that level will be about 44 km sec^{-1}. If Chandrasekhar's fundamental assumption, that the chromospheric material is supported on the average by radiation pressure, could be admitted, turbulent motion would certainly result.

Chandrasekhar refines the problem by assuming the sun to be dotted with a series of alternate bright and dark regions. Simplification is effected by treating only the two-dimensional case, motion in the vertical and horizontal planes, and by assuming a simple sinusoidal law for the intensities of the radiation. He thus imagines the sun to be covered with a regular grill work, with perfectly regular spacing. The orbits of atoms subject to radiation pressure from this non-uniform source have been worked out by Chandrasekhar. He lays emphasis on the fact that certain of these orbits are periodic and he restricts the problem to orbits of this type. He points out that the chromospheric boundary, defined as the envelope of these trajectories, should present a wavy outline, rising over the hot areas and descending over the cool ones. If the chromosphere were actually built upon a geometric pattern with a regular spacing of hot-spots, this mathematical procedure would be unobjectionable. Stray atoms, moving in aperiodic orbits might arise, but the ones in periodic orbits, favored by resonance, would predominate.

No such geometric pattern exists, however. Individual atomic trajectories are controlled by the radiation field the atom happens to be moving through and the particle could not possibly adjust its orbit to fit the conditions demanded by the periodic solution. Atoms would

tend to fall short or overshoot the mark and the chromosphere would have to be built anew at each hot-spot. Unfortunate as it is for the theory, it is fortunate for observation that such is the case. The chromosphere does not *look* like a wavy sea. It looks more like an ocean dotted with water-spouts.

The observational data on the motions of solar prominences as exhibited by the McMath motion pictures and interpreted by Mc Math and Pettit show that the motions are extremely complicated. On occasion prominences appear to act in the classically expected fashion, with eruptions and geyser activity. Sometimes the prominence material rains down from an apparent source high above the chromosphere. In many instances there is no obvious replenishment of the source. Sometimes the motions are violent, suggestive of a lightning flash. Sometimes the great clouds of gas remain suspended at considerable heights and remain for days, almost unchanged in form. On occasion, spot activity and prominence activity are seen to be closely associated.

The McMath records seem to confirm some earlier observations about the motions of gases in prominences. The most important observation has been somewhat inexactly referred to as Pettit's first law of prominence motion. When the distance, s, through which any small knot of prominence matter has moved from some arbitrary starting point, is plotted as ordinate against the time as abscissa, the resulting plot appears to be made up of a series of connected straight lines of different slopes. The interpretation is that the velocity of the knot remains constant for a considerable length of time, is then subject to a short and sudden acceleration, after which the velocity is steady again, etc. When the path is curved, s is measured along the trajectory. The law appears applicable to either descending or ascending material. It fits most accurately when the prominence is eruptive in character, with the motion nearly perpendicular to the solar surface. For the highly curved trajectories, the effect of projection introduces an unknown distortion that cannot be entirely eliminated.

Although the observational evidence appears to be in favor of the reality of the first law of motion, the second law of motion may perhaps be questioned. Pettit has argued that the new velocity, after one of the sudden accelerations, is an integral multiple either of the preceding velocity or of the next earlier velocity. More observational evidence must be obtained before the second law, which involves a peculiar quantization effect, can be accepted without reservation.

In addition to the Pettit laws, a few further statements can be made concerning the character of the atmospheric motion in the chromosphere

and prominences. Many prominences show a tendency to arch over and connect with other prominences. From the great cap thus formed, there extend "tubes" or "feeders" connecting the cap with the lower chromosphere. Presumably material is either ascending or descending in the tube. At a considerable altitude above the photosphere the prominence material appears to acquire a velocity component parallel to the solar surface. This motion results in the typical arch formation presented by many prominences. There is one further impression that one gains from viewing the McMath and Lyot records. The paths or trajectories of prominence material appear to be conserved, once they are established. In other words, if prominence material has moved along a given path, there will be a tendency for other material to follow along in approximately the same path even though a considerable time has elapsed. Further observation is required to substantiate the effect.

The appearance of the prominences is so striking that some scientists have concluded that the effects may be due to waves of excitation rather than to actual motion of matter. The proposed explanation would make the prominences a phenomenon not dissimilar to the Aurora Borealis. The observations, however, immediately show that this explanation is untenable. The observed radial velocities and Doppler shifts prove that the velocities are real.

Several general conclusions may be drawn concerning the nature of forces acting. The following types of forces may conceivably be operative. The fundamental force, of course, is gravitation. This downward acceleration may be balanced, in part, by the gradient of gas pressure in the corona and in the prominences. The pressure of radiation on the collisions of electrons or other material particles emitted from the solar surface also may tend to counter-balance gravitation. In addition, the effect of the electric and magnetic fields must be considered:

Pettit's first law of motion, apart from the sudden acceleration, shows that the forces tend to balance out along the line tangent to the path. The curvature of the path, then, must arise from a force acting normal to the trajectory, whereas the sudden acceleration, unaccompanied by a change in direction, must be attributed to a force acting along the trajectory. Pettit refers to that region of the chromosphere toward which several prominence streamers may tend to converge as "a center of attraction". The converging trajectories suggest attraction, but the failure of the prominence material to show any acceleration along its path indicates that the word "attraction" should not have been used.

One of the earliest theories of prominence activity was that worked out by Milne, who based his investigation on his earlier theory of chromospheric equilibrium. Milne assumed that radiation pressure was the force acting to oppose gravity. Consider a motionless atom, absorbing radiation in the center of an absorption line and just in balance between radiation pressure and gravitation. If this atom were suddenly to acquire an outward component of velocity, the resultant Doppler effect would cause the atom to absorb in the line wing where it would be subjected to more intense radiation. In consequence its equilibrium would be disturbed. Acceleration would occur and the atom would eventually be blown away from the sun.

Milne calculated the trajectories and velocities of matter acting under the conditions described, for an absorption line with a V-shaped profile (Fig. 1). More recently, Pettit and McMath have extended the

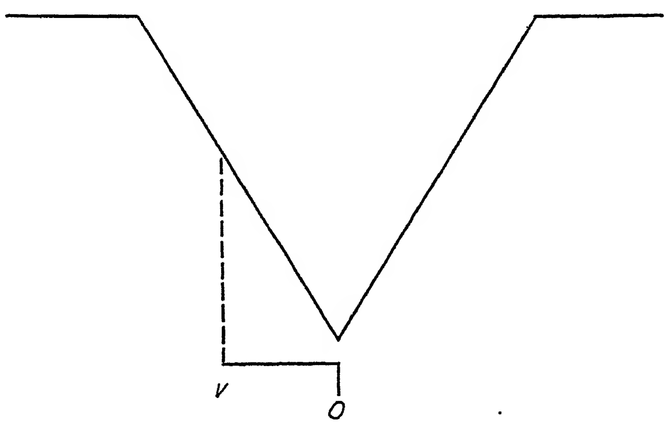


Fig. 1. Milne's Theory. Atom moving outward absorbs in violet wing (v) of absorption line.

calculations to the case of a U-shaped line as well. In neither example do the theoretical velocities show any correspondence with the observed. An analogous situation exists in certain trajectories calculated by Bobrovnikoff.

The constancy of the observed velocities requires the very special physical condition that the accelerations be zero. On the hypothesis that radiation pressure is responsible for the motion of prominences, a very delicate balancing is required. My first attempts to account for the recovery of balance, after a sudden increase of radiation pressure had led to acceleration, were based on the assumption that the greater amount of radiation would lead to further expulsion of atoms. The increased material, in turn, was expected to reduce the flux to its original value, so that radiation and gravity would again balance.

Two difficulties were experienced, however, in developing the theory outlined in the preceding paragraph. One lay in finding how the newly ejected material would acquire the proper velocity just to screen the overlying layers. This difficulty is, perhaps, not insuperable, but the failure of the observations to disclose the presence of any newly ejected sheet at the moment of the observed acceleration constitutes a serious objection to any theory based on this variety of equilibrium.

What is needed to account for the observed motion is a positive mechanical action that will provide either acceleration or deceleration, for atoms moving too slowly or too fast, respectively. A mechanism behaving in this fashion will lead to stability of motion. Fortunately there appears to be one definite method for obtaining stable equilibrium, under the action of radiation pressure. A significant feature of the theory is that it leads directly to a prediction of the constant-velocity law.

We start with a very simple hypothesis. In effect we reverse the conditions of Milne's theory. Assume the line to be emission, instead of absorption, and that an atom absorbing at or near the center of the line is exposed to radiation pressure greater than the downward gravitational acceleration. We suppose, also, that gravity will exceed radiation pressure for an atom absorbing in the far violet wing of the emission line, because of the high upward velocity. At some intermediate upward velocity, the forces of gravitation and radiation pressure will exactly balance and the atom will be subjected to either a downward or upward force until this equilibrium is attained.

The elementary theory can be thrown into simple mathematical form. Consider an emission line with a profile as depicted in Fig. 2,

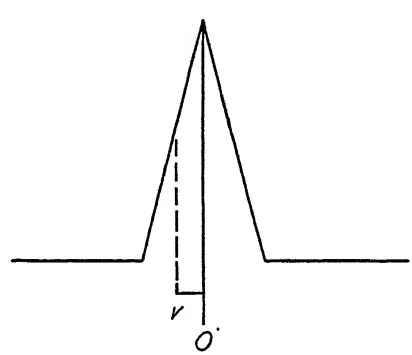


FIG. 2. Outwardly moving atom absorbs in violet wing (ν) of emission line.

triangular in shape with central flux F_0 at $\nu = \nu_0$ and half-width b expressed in frequency units. The flux, then, at any frequency in the

violet wing will be

$$F = F_0\left(1 - \frac{\nu - \nu_0}{b}\right). \tag{2}$$

The radiation pressure exerted on an atom is

$$p_r = F\frac{\pi\epsilon^2}{mc}f, \tag{3}$$

where $\pi\epsilon^2 f/mc$ is the integrated absorption coefficient. Further, if v is the upward velocity component, we have

$$\frac{v}{c} = \frac{\nu - \nu_0}{\nu_0}. \tag{4}$$

Hence

$$p_r = F_0\left(1 - \frac{\nu_0 v}{bc}\right)\frac{\pi\epsilon^2}{mc}f. \tag{5}$$

The gravitational force is $m_a g$, where g is the acceleration and m_a the atomic mass. The equation of motion becomes:

$$m_a\frac{d^2r}{dt^2} = F_0\left(1 - \frac{\nu_0}{bc}\frac{dr}{dt}\right)\frac{\pi\epsilon^2}{mc}f - m_a g. \tag{6}$$

In writing (6) we have ignored inverse-square effects.

The first integral of (6) is easily obtained from the relation

$$\frac{dr}{dt} = v. \tag{7}$$

Thus:

$$v = \frac{\mu - 1}{\mu}\frac{bc}{\nu_0}\left(1 - e^{-\frac{\mu g \nu_0}{bc}t}\right), \tag{8}$$

where the constant of integration is determined so that $v = 0$ for $t = 0$. We have set μ equal to the ratio between radiation pressure and gravity at the center of the line, or:

$$\mu = \frac{F_0}{m_a g}\frac{\pi\epsilon^2}{mc}f. \tag{9}$$

For the physical conditions of (8) to apply, we must have $\mu > 1$. The limiting velocity, v', is

$$v' = \frac{\mu - 1}{\mu}\frac{bc}{\nu_0}. \tag{10}$$

Equation (8) may be written

$$v = v'(1 - e^{-t/\tau}), \tag{11}$$

where τ is the mean relaxation time:

$$\tau = \frac{v'}{(\mu - 1)g} \cdot \tag{12}$$

From the observational data of McMath and Pettit, we conclude that τ is of the order of a few minutes. v' ranges from 10 to perhaps 100 km sec^{-1}. Adopting the latter figure, and taking $g = 2.7 \times 10^4$, we have

$$\mu \sim 4. \tag{13}$$

The precise numerical magnitude is of no significance. But if the physical processes here discussed are to be seriously considered, we see that a strong overbalance of radiation pressure at the line center is required.

A single emission line will account for constant upward velocities. The occurrence of constant downward velocities requires somewhat more complicated physical conditions. An emission line of the form shown in Fig. 3, however, will enable constant velocities of descent to be

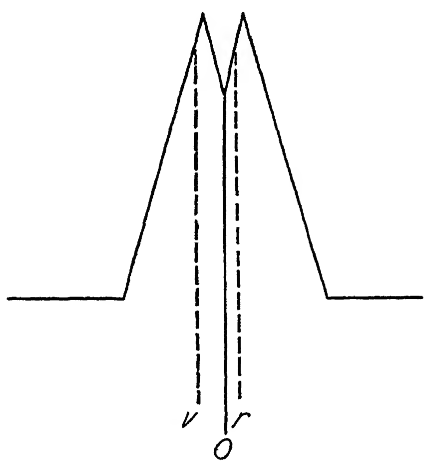


FIG. 3. Self-reversed emission line permits either outwardly (v) or inwardly (r) moving atoms to attain balance against gravitation.

obtained. Emission lines of this sort are common in both laboratory and stellar spectra. Thus profiles are attributed to the phenomenon of self-absorption. It is interesting to note that the K_3 line of ionized calcium shows a similar profile in the solar spectrum.

The amount of mass to be supported against the force of gravitation is quite large. Gas pressures may amount to 10^{-8} or 10^{-7} atmospheres. Two independent estimates of the amount of material to be supported (chiefly hydrogen), one from the observed intensity of lines of the flash and prominence spectra at time of total eclipse, and the other from the

appearance of prominences in absorption in spectroheliograms of the solar disc, agree in showing that the material to be supported in a fair-sized prominence is not less than 10^{-4} grams per cm^2.

To this prominence material gravity communicates momentum at the rate of about 3 gm cm sec^{-2} per cm^2. For this downward momentum to be compensated by radiation pressure, the sun must emit the appropriate amount of energy in the Lyman region. Radiation lying to the red of the Lyman limit, except that within the Lyman lines, cannot contribute appreciably to the support of hydrogen atoms, because the majority of the atoms are in the ground level. Further, the optical depths in the Lyman lines and continuum must be sufficient to enable the atmospheric atoms to absorb an appreciable amount of the emitted radiation. Analysis shows that, if radiation pressure is the source of chromospheric support, the effective temperature for the ultra-violet radiation at the base of the chromosphere must be of the order of twenty thousand degrees *K*.

One of the difficult problems connected with the radiation-pressure explanation, is the previously mentioned tendency of prominence material to acquire horizontal velocity at high levels. It is conceivable that part of the effects may be attributed to intense Lyman radiation emitted by the body of the prominence. Or possibly the effect may arise from an asymmetrical distribution of "hot-spot" sources of ultra-violet energy. But the curvature of the filaments, and their apparent tendency to follow the magnetic force fields of sun-spots, suggest that other forces may be active.

We know, of course, that sun-spots possess intense associated magnetic fields, which may reach on occasion the value of 4,000 Gauss. In addition, Hale has found that the sun possesses a general magnetic field of the order of 50 Gauss. The magnitude and character of the general solar magnetic field have been derived on very slender observational evidence. The magnitude of the field is probably established beyond question, but certain other characteristics, such as the small inclination between the magnetic axis and the axis of rotation, the reported precessional rotation of the former about the latter axis, and finally the so-called "radial limitation" of the magnetic field must still be regarded with suspicion. The arguments for radial limitation are open to serious question. The reasoning is based, in part, on what appears to be an erroneous interpretation of the significance of chromospheric "heights." Also, it is questionable whether a proper allowance has been made for the relative strengths of the various absorption lines in the interpretation of the Zeeman Effect. Chapman and Gunn have

attempted to derive a theory of the radial limitation of the field. The argument centers about crossed electric and magnetic fields, with electrons spiralling around the magnetic lines of force. Cowling and Ferraro, in more detailed analyses, have concluded that, although radial limitation may possibly be explained on the basis of certain rather artificial assumptions, there are nevertheless many difficulties with the theory. There is no apparent mechanism for maintaining the intense electric field.

Further evidence in favor of a solar magnetic field has been obtained by Le Maitre and Vallarta, who have concerned themselves with the effect that a solar magnetic field would exert on the cosmic ray trajectories. They have found that a field of the order of 30 Gauss will explain the observed cosmic ray effect. Dr. O. Godart, Belgian Foundation Fellow, has kindly communicated to me certain unpublished results concerning the diurnal cosmic-ray variation, which also seem to lead to a solar magnetic field of the same order of magnitude.

Thus far, no satisfactory theory for the origin of the magnetic fields either of the sun itself or the sun-spots has been proposed. Nevertheless, in one sense there is no mystery about the origin of these fields. Since they cannot possibly be associated with permanent magnetism, we must attribute them to electric currents of some variety. It is easy to show that no intense electrical fields can exist on the sun, hence the currents that give rise to the magnetic fields must be galvanic in character. To maintain a field of 4,000 Gauss over the area of a large sun-spot requires currents of the other of 10^{14} amperes.

The general appearance of solar prominences suggests that their form, if not their motive power, may be controlled by an associated magnetic field. It is uncertain under what conditions magnetic fields can arise in the course of circulation of an ionized but macroscopically neutral gas. But the indications are fairly clear that prominences, like sun-spots, may possess a magnetic field. Furthermore, magnetic poles do not occur singly. Each positive pole must be accompanied by its negative counterpart, resulting in a "dipole." The arching effect of prominences previously described is not dissimilar to the lines of force surrounding two neighboring bar magnets. Careful observations of the Zeeman effect for low-lying prominence material should be carried out.

The presence of a magnetic field around which the ionized atmospheric gases could spiral would explain one of the most complicated of the observed chromospheric phenomena, viz., the occasional tendency of large clouds of gases to become condensed into smaller volume during descent.

It is entirely within the realm of possibility that some electrical fields may exist in the solar atmosphere. As Lindemann showed many years ago, however, the sun could acquire through loss of the volatile electron gas, a maximum positive potential of some 2,000 volts. Any higher potential would lead to direct expulsion of positive electrons against the downward pull of gravitation. The potential field of 2,000 volts could be set up by removing one electron from every six cm^2 of solar surface. It is easy to show that the various masses of gas that comprise the solar atmosphere are very nearly neutral electrically. Even so, the action of radiation pressure on an ion moving through a magnetic field under the influence of the combined forces of gravitation and radiation pressure, would tend to deflect the ion at right angles to both force fields. The accompanying electrons would be deflected in parallel paths but in opposite direction to the ions. In consequence, masses of gas suspended in the solar atmosphere would tend to become electrically polarized. As the electrical circuit may possibly be completed through the lines of magnetic force in the solar interior, a galvanic current may flow through the ionized gases of the clouds.

Surprise is often registered that the forms of prominences as recorded in the light of ionized calcium and in the light of neutral hydrogen are so similar. The implication is that, because neutral atoms should not respond to a magnetic or electric field, the distribution of hydrogen should be different from that of calcium. This argument is not valid, however. The solar observations indicate that hydrogen exists in the ionized form a large percentage of the time. Consequently, the fact that it is recorded spectroscopically only when it has picked up an electron and has thereby been rendered neutral is incidental. The problem of helium is a little more difficult. The great intensity of the line 4686 of ionized helium in the flash and prominence spectra would suggest, however, that even helium is very appreciably ionized. Excitation temperatures in the ultra-violet from twenty to twenty-five thousand degrees are required for this condition to be fulfilled. There appears to be one significant difference between the distribution of helium in a prominence and the distribution of other materials. Helium shows a definite tendency to become more intense at high levels. It remains to be seen whether or not this difference is to be attributed to an effect of the low atomic mass or to the conditions of excitation.

One of the greatest unsolved problems of the prominences and chromosphere is the question of the so-called "coronal" prominences. These are the prominences that appear to stream from some source high above the photosphere. There is no visible mode of replenishment for

the source of the streamers. It is tempting to suggest that the material is condensing from the coronal gases, and such may indeed be the case. For this explanation to be acceptable, however, some very special conditions must be fulfilled. The corona itself appears not to be a source of hydrogen radiation. If hydrogen be present at all, therefore, and that is the assumption we are examining, it must exist in some condition of temperature and pressure where the ordinary Balmer spectrum is suppressed. There is no normal way in which this can be accomplished, except under the conditions of extremely low pressure. If, through the action of magnetic fields or otherwise, the coronal material be allowed to stream into the field and thereby be compressed, one may possibly find an explanation for the phenomenon.

Much work remains to be done, both observationally and theoretically, before anything like a complete picture of prominence activity can be drawn. Further studies of the relationship between the corona and prominences are greatly to be desired. One should realize that observations indicate an electron pressure in the corona not sensibly different from that found in the chromosphere. Differential pressure effects and possible temperature gradients introduce complications additional to those of the fields of force already enumerated. In spite of the difficulties, however, two points emerge from the foregoing discussion. It is possible for radiation pressure to be self-compensated as a field of force. The radiation required for the support and for the excitation of the flash and prominence spectra agree in order of magnitude. Finally, the assumption that magnetic fields are associated with prominences as well as with sun-spots appears to clear up additional effects associated with the problems of prominence form and motion.

THE COMPOSITION OF THE SUN

CHARLOTTE M. SITTERLY

Princeton University Observatory

(*Read February 17, 1939, in Symposium on Progress in Astrophysics*)

THE spectroscope has yielded a wealth of information about the composition of the sun. As early as 1895 Rowland succeeded in photographing the solar spectrum from λ2975 to λ7330 and identified some 39 elements in the sun.[1] In this classical piece of work accurate positions and estimated intensities of some 20,000 lines are recorded. Rowland's photographs were remarkable for their fine detail; many of the faintest lines have never been observed elsewhere even to-day. The measures, made by Jewell, are of exceptional precision and the entire work is unusually free from error.

The solar spectrum exhibits spectral lines of many elements. The oldest and most obvious method of determining the chemical origin of solar lines is that of coincidence, *i.e.* photographing the spectrum of any element and comparing it directly with a photograph of the solar spectrum. Each element has its own characteristic pattern or set of spectral lines. In the ideal case every laboratory line can be matched in the solar spectrum, as regards intensity and position. Of all the elements, iron presents the most convincing evidence of this sort. In fact it is difficult to produce a laboratory arc or spark spectrum of iron showing as many iron lines as are present in the sun. Other elements such as calcium, magnesium, silicon, aluminium and sodium can be identified without question. All the principal lines of titanium and vanadium are present. Only the stronger lines of copper and zinc occur but these are easily detectable; and for silver the two strongest lines are weak in the sun, and others absent.

Rowland's early work on the solar spectrum is sufficient to provide a good general knowledge of the composition of the sun. He listed 39 elements as present, all but 2 of which have been confirmed by later observations. The great differences in the behavior of various elements, as regards intensity in the sun and number of lines present, immediately suggest that they are not equally abundant.

[1] *Ap. J.*, 1–5, 1895–97.

In 1928 a revision of Rowland's Preliminary Table of Solar Spectrum Wave-lengths [1] was made at Mount Wilson by Charles E. St. John and others. The identifications made by Rowland were greatly extended but this problem is still far from complete.

It has long been desirable to extend the observations of the solar spectrum to the infra-red. In 1919 Meggers reached the limit λ9000.[2] A recent notable piece of research has been done by Babcock at Mount Wilson. Owing to the development of red-sensitive plates, he has succeeded in photographing the spectrum to λ13,500, adding more than 4700 lines to Rowland's original list. Measures of the last thousand Ångstroms in the extreme red are not yet completed. Because of the decrease of sensitivity of Rowland's plates in the red, Babcock is also observing the overlapping region from λ6600–λ7330, where several hundred new but faint lines have been recorded. Although numerous band lines due to oxygen and water vapor in the earth's atmosphere mask the solar spectrum in certain parts of the red, yet a number of important atomic lines are present, some of the strongest being due to carbon, oxygen, sulphur, and silicon. The accuracy of Babcock's measures far surpasses that of laboratory measures in the red and in fact led to the discovery of the presence of two series of magnesium lines which have been heretofore very incompletely observed.[3] Phosphorus was also detected as a result of this work.[4] Lines of the non-metals, in general, appear in the infra-red, *i.e.* beyond the limit of Rowland's plates, which explains why he did not find them.

Elements may be classed in several grades, varying from those with nearly every line present to those the strongest lines of which are mere traces in the solar spectrum. It would be wrong, however, to assume that because an element is represented in the sun only by weak lines, it must necessarily be rare. A band of ozone in our atmosphere masks the solar spectrum to the violet of λ2975 and the leading lines of several elements lie in the masked region. This is true of almost all of the non-metals except hydrogen.

Various factors must be considered in determining the chemical origin of lines in the solar spectrum. The intensities of both solar and laboratory lines must be taken into account, for chance coincidence may easily lead to spurious identifications. If the weaker lines of an element agree in wave-length with solar lines, but the strongest lines are absent, the agreement is accidental.

[1] *Carnegie Publ. No. 396; Papers of the Mt. Wilson Observatory*, III, 1928.
[2] *Publ. Allegheny Observatory*, **6**, 13 (No. 3), 1929.
[3] Russell, Babcock and Moore, *Phys. Rev.*, **46**, 826 (No. 9), 1934.
[4] Moore, Babcock and Kiess, *Ap. J.*, **80**, 59, 1934.

Lines of different elements or two lines of the same element frequently contribute to the production of one solar line; consequently the solar intensity and wave-length are those of a blend. A strong line of one element may also completely mask another line. The question of blends and masked lines makes the detection of some elements in the sun rather difficult—as will be seen later.

A quantitative as well as a qualitative analysis of the solar spectrum can be made by a careful study of the identification of solar lines. It is obvious that hydrogen, sodium, magnesium, silicon and iron are abundant; titanium, vanadium and copper less abundant; while silver is relatively rare.

Fortunately, the spectra of different portions of the sun can be studied. So far only the spectrum of the solar disk has been mentioned, but the spectra of sun spots [1] and of the solar chromosphere [2] have also been extensively observed. In these spectra hundreds of lines have intensities noticeably different from those of the solar spectrum, although their measured positions are in agreement.

Both Lockyer and Fowler [3] called attention to the fact that lines intensified in passing from the disk to the spot spectrum were produced at low temperature in the laboratory; while those intensified in going from the spectrum of the electric arc to that of the electric spark behaved in the opposite manner in the spot. Likewise, Hale, Adams and Gale [4] realized that a difference in temperature between disk and spot would explain the behavior of certain elements. Their laboratory experiments initiated the long program of temperature classifications of spectral lines by King at Mount Wilson. The "low temperature" lines, those produced with the least excitation and most persistent, are called the "Ultimate Lines." Thus King grades the lines appearing at increasing temperatures of the electric furnace and arc in the classes I to V.

The significance of all this has been made clear by theoretical investigations on the origin of spectra. The rapid progress in this field has greatly aided the astrophysicist in disentangling solar and stellar spectra, and has furnished him with more refinements than are possible by the "coincidence" method. Generally speaking, the ultimate lines are absorbed by atoms in their lowest energy states. More excitation will produce lines absorbed by atoms in higher states, *i.e.* in which they

[1] *Mt. Wilson Contr. No. 446; Ap. J.*, **75**, 222, 298, 1932.
[2] Mitchell, *Ap. J.*, **71**, 1, 1930; Menzel, *Publ. Lick Obs.*, **17**, Part 1, 1931.
[3] *Chemistry of the Sun* (1887), p. 314; *Trans. I. A. U.*, **1**, 201, 1906.
[4] *Mt. Wilson Contr. No. 5; Ap. J.*, **23**, 11, 1906. *Mt. Wilson Contr. No. 11; Ap. J.*, **24**, 185, 1906.

have more energy. Still higher energy will excite the atoms to such a degree that they lose electrons. This will result in the formation of an entirely new spectrum, namely the spark spectrum. The lines are then "enhanced," the term enhanced meaning that they are increased in intensity in the spark spectrum as compared with the arc spectrum.

Given the measured wave-lengths, estimated intensities, and temperature classes of the lines in a spectrum, the physicist analyzes the spectrum, handling each arc and spark spectrum separately. Every spectral line is produced by a transition between two energy levels. Such an analysis reveals closely related groups of energy levels which form spectroscopic terms; and transitions between these terms produce groups of lines called multiplets. The lines of a multiplet behave as a unit; the relative intensities remain substantially the same, except under very unusual circumstances.

An absorption line can be produced only when the atom is already excited to the lower energy state of the transition in question. The *excitation potential* for any given line may here be defined as the energy expressed in electron volts, required to raise the atom from its lowest energy state to the state in which it can absorb the line. For ultimate lines this quantity is zero or very small. The *ionization potential*, similarly, measures the energy required to ionize the atom. It is an extremely important factor in the interpretation of all astronomical spectra.

In searching for an element in the sun, the *leading* lines of those multiplets containing the *ultimate* lines of the element should be present, *i.e.* the strongest related lines of lowest excitation potential. If these lines are accessible but absent, the element is not present. Each multiplet should also be considered as a whole, *i.e.* in any multiplet the relative intensities of the lines in sun and laboratory should be consistent. In this connection, blends and masked lines must receive careful consideration.

The ultimate lines of many elements lie to the violet of $\lambda 2975$ and cannot, therefore, be observed in the solar spectrum. However, if an element is present in sufficient quantity in the sun, it is represented not only by ultimate lines but also by those of higher excitation. The search should then be made among the strongest lines of multiplets of lowest excitation potential in the accessible region. If the leading lines of these multiplets are absent, the coincidence of wave-length of weaker members is accidental.

The differences in behavior of lines in spot, disk and chromospheric spectra are analogous to those in laboratory spectra produced in the

electric furnace, arc and spark respectively. It has already been suggested that a difference of temperature would account for the strengthening of the ultimate lines of some elements such as calcium, scandium, titanium and vanadium in going from the disk to the spot spectrum, just as is the case in comparing arc and furnace spectra of these elements. It is now known that the temperature of the spot is some thousand degrees cooler than that of the disk and thus the earlier explanation is substantiated.

Elements of low ionization potential are easy to ionize in the laboratory. Similarly, in going from spot to disk spectra the higher temperature of the latter causes such elements to become completely or almost completely ionized. This is well illustrated by the behavior of the ultimate lines of lithium, rubidium and indium. These elements all have low ionization potentials. For the first two, the lines are extremely faint in the disk spectrum and greatly strengthened in the spot. The third, indium, is found only in the spot spectrum.

Most of the rare earths appear only in the ionized state in both disk and spot spectra, since they have low ionization potentials. The two exceptions are ytterbium [1] and europium.[2] The arc spectra of these elements have such a structure that the energy, which in other rare earth spectra produces hundreds of lines, is concentrated in the production of a few very conspicuous ultimate lines. For ytterbium, one of the strongest lines is blended in the sun, and the other is faintly present in the disk spectrum and slightly strengthened in the spot. Of the three outstanding arc lines of europium, one is masked, and two are present only in the spot spectrum.

More energy is required to produce arc lines of high excitation potential, and lines due to ionized elements. Hence these lines are *stronger* in the disk spectrum than in the spot, owing to the higher temperature of the disk.

The behavior of compounds in sun and spot spectra may also be mentioned. In the disk spectrum greater excitation due to the higher temperature dissociates most molecules. Lines of titanium oxide and calcium hydride appear only in the spot spectrum—a fact due to dissociation of the molecules as well as to ionization of the atoms.

A detailed critical study of elements in the sun, made during the last few years, is summarized in Table I. Here the elements are divided into four groups: present, insufficient solar data, insufficient laboratory data, and absent.

[1] Moore and Meggers, *Publ. A. S. P.*, 44, 175, 1932.
[2] Russell and King, *Mt. Wilson Contr. No. 608; Ap. J.*, 89, 384, 1939.

Of the 64 elements listed as present, a few deserve special mention. Helium was first discovered in the flash spectrum at the total eclipse of 1868, but was not isolated in the laboratory until years later.[1]

Boron has no lines in the observable region of the solar spectrum and its presence is detected only in a compound with oxygen.[2] Fluorine is similarly identified by a compound with silicon.

TABLE I

ELEMENTS IN THE SUN

Present	64	H He Li Be B C N O F Na Mg Al Si P S K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge Rb Sr Y Zr Cb Mo Ru Rh Pd Ag Cd In (Sn) Sb Ba La Ce Pr Nd Sa Eu Gd (Tb) Dy Er Tu Yb Lu Hf (Ta) W (Os) (Ir) Pt Pb
Insufficient solar data	2	Ne Cs
Insufficient laboratory data	7	Ma Il Ho 85 87 Th U
Absent	19:	
Ultimate lines accessible	3	0.0 0.0 0.0 0.0 0.0 Cs* Re Tl Bi Ra*
Ultimate lines inaccessible	11	2.3 4.6 4.9 5.5 6.0 6.9 As Au Hg Te Se I 7.8 8.3 8.9 9.9 11.5 16.6 Br Xe Cl Kr A Ne*
Not to be expected	5	Po Rn Ra Ac Pa

* Entered twice.

The evidence of the presence in the sun of the elements in parentheses in the Table, *Sn, Tb, Ta, Os* and *Ir,* is not so strong as for the other elements, for various reasons. Several of these have been discussed in detail in a previous publication and no new data are available.[3] Others have been revised.

The detection of *Sn* has presented an interesting problem. The leading accessible line occurs in the arc spectrum of iron. Only recently has it been determined that the line is due to tin alone, which appears as an impurity in the iron used to produce the laboratory spectrum. Two other strong low level lines of tin are masked in the sun and the third, if present, is probably a blend. Two of these four lines lie to the violet of λ3050, close to the region where the earth's atmosphere becomes

[1] Russell, Dugan and Stewart, *Astronomy*, **2**, 504.

[2] Nicholson and Perakis, *Mt. Wilson Contr. No. 370; Ap. J.*, **68**, 327, 1928. Russell, *Mt. Wilson Contr. No. 490; Ap. J.*, **79**, 317, 1934.

[3] Moore, *Mt. Wilson Contr. No. 565; Ap. J.*, **85**, 79, 1937.

opaque. There is independent evidence that the observations here are very difficult. For example, two strong laboratory lines of silicon at $\lambda 2970$ and $\lambda 2987$, which should be present, were not recorded by Rowland.

Terbium is detected almost entirely by its presence in the flash spectrum. It occurs only in the ionized state, as is to be expected, but several of the leading lines are badly disturbed by blends. More laboratory material is needed for this element.

The three low level lines of tantalum to be expected, if the element is present in the solar spectrum, agree with very faint lines recorded by Rowland. There is some question as to the reality of two of these solar lines, as it is possible they may be "ghosts,"—lines which are caused by a special type of defect in the grating.

Most of the leading accessible lines of osmium and iridium are either masked or blended in the solar spectrum. Probably no further observational material will clarify the question of the presence of these elements.

In spite of the complications concerning these five elements, *Sn*, *Tb*, *Ta*, *Os* and *Ir*, there is reasonable evidence to call them all present.

The two elements *Cs* and *Ne* are classed under the heading, "Insufficient solar data." Adequate observations of the spot spectrum in the infra-red will settle the question of the presence of the ultimate lines of caesium.[1] More observations of the spectrum of the chromosphere are necessary in order to determine whether neon is present or absent.

Seven elements must be listed in the class, "Insufficient laboratory data." Some laboratory work has been done fairly recently on thorium and holmium, but the analysis of these spectra is needed in order to decide whether they are present in the sun. It should be possible to secure good laboratory data for *Ho*, *Th* and *U* within a reasonable length of time. Nothing is known at present about the spectra of *Ma*, *Il*, 85 and 87, for the very good reason that they have not been isolated in sufficient quantity. In fact the chemists debate whether they have been isolated at all.

In Table I, 19 elements are listed as absent. These are divided into three groups: "Ultimate lines accessible," "Ultimate lines inaccessible," and "Not to be expected." In the first two groups the excitation potential of the accessible lines is entered above the chemical symbol.

Those in the last group are radioactive and of relatively short life.

The high excitation potentials of the accessible lines, in the second group of absent elements, are hopelessly unfavorable for their detection.

[1] A letter from Mr. Babcock states that *Cs* is absent (Feb. 1939).

Russell has shown conclusively [1] that "the principal factor which is unfavorable to the appearance of a spectral line in the sun is a high excitation potential." There are few solar lines for which this exceeds five volts; the only very conspicuous ones being due to hydrogen, which is known to be extremely abundant. An increase of one electron volt makes the appearance of a line in the sun less likely by a factor of 10. For *As* this factor is not so serious, since the lines are intersystem combinations and but few of the excited atoms would absorb them.

The remaining three elements, *Re*, *Tl* and *Bi*, could certainly be detected if they were present in quantities comparable with any of those which have been found. There can be no doubt that if these are present they are extremely rare. All three are among the rarest elements on earth on the basis of percentage in the accessible crust. Rowland expressed doubt as to the presence of *Bi*. His identifications of *Tl* result from accidental coincidences. The leading ultimate *Tl* line observable in the sun-spot spectrum coincides in wave-length with a faint solar line which is not strengthened in spots. Consequently it is not due to *Tl*, which is easily ionized. These are the only two elements removed from Rowland's list.

A quantitative determination of the abundance of elements in the sun can be made from the estimated and measured intensities of their lines. The number of atoms producing a line of given intensity can be calculated. The amount of ionization of different elements must also be taken into account. Allowing for this factor the number of neutral or ionized atoms per unit area in the solar atmosphere can be determined for any element, provided the excitation potentials of the lines present in the sun are known.

In this way Russell [2] has calculated the abundances in the sun's atmosphere of 63 elements. From a comparison with Goldschmidt's [3] investigations in geochemistry he concludes that "The general composition of the earth and sun is similar except that hydrogen and helium are enormously more abundant in the sun. The relative proportions of the metals are remarkably alike. Recent studies emphasize the similarity. For example scandium, once supposed to be much rarer on earth than in the sun, has been detected spectroscopically in many types of rock." As more becomes known the agreement in composition of earth, sun and meteorites becomes better and better. The probable composition of the sun's atmosphere as quoted from Russell [4] is given in Table II.

[1] *Mt. Wilson Contr. No. 383; Ap. J.*, **70**, 11, 1929.
[2] Russell, Dugan and Stewart, *Astronomy*, **2**, 503.
[3] *Nach. der Gesellschaft der Wiss. zu Gottingen—Math-Phys. Klasse*, 1931–35, and others.
[4] *Mt. Wilson Contr. No. 383*, 63; *Ap. J.*, **70**, 73, 1929.

TABLE II

Element	By Volume	By Weight
Hydrogen	60 parts	60
Helium	2?	8?
Oxygen	2	32
Metals	1	32
Free electrons	0.8	0
Total	65.8	132

Although a comprehensive study of the sun's composition has been made in recent years, the work is far from complete. As the analysis of laboratory spectra progresses, more material becomes available for the identification of solar lines, and many revisions of the "Revised Rowland Table" have already been made. This is particularly true of the rare earths, the identifications of which in the sun need extensive revision.

It is highly desirable to secure new photographs of the solar spectrum near the violet limit set by our atmosphere; and of the spot spectrum over the entire visible range. Fortunately the sun-spot maximum just passed has enabled Babcock to obtain some splendid spot spectra. One of particular interest is in a region undisturbed by atmospheric lines, near $\lambda 10,500$, where there are strong lines of *Si*, *S*, *Ti*, *Fe* and of other elements. A study of these photographs is now in progress.

When the large publication on the Infra-Red Solar Spectrum is completed, and when more measured intensities of solar lines are available it will be desirable to recalculate the abundance of elements in the sun. In particular, the values for carbon, silicon and magnesium need revision and others such as phosphorus and sulphur can be improved.

The question of the abundance of hydrogen has a most important bearing on the entire subject of stellar constitution. Many other aspects of solar work might also be mentioned. It is hoped that the present brief survey has revealed some of the interesting problems presented by a study of the composition of the sun.

THE CONSTITUTION OF THE PLANETS

RUPERT WILDT

Princeton University Observatory

(*Read February 17, 1939, in Symposium on Progress in Astrophysics*)

DURING the past decades most impressive progress has been made in analyzing the physics of the atoms and of the stars, and the intimate relations between the microcosm and the macrocosm of modern science have often been commented upon. It is on account of the comparatively simple behavior of matter at high temperatures and in the state of atomic dispersion that the theory of the constitution of gaseous stars has already reached a remarkable degree of perfection and is capable of being expounded in a largely deductive manner. No such course can be followed in discussing the current views concerning planetary constitution. They have evolved by convergence of several lines of approach, among which might be mentioned the geodetical one, the geophysical and the geochemical one. As these terms indicate, the study of the Earth has furnished the pattern following which the features of the planets are interpreted. It is of interest to note that the ideas and conceptions used in this interpretation are truly of old, in contrast to the principles now recognized as explaining the constitution of the stars. This point is nicely illustrated in a posthumous work by Christian Huygens, to whom 17th century astronomy owes, besides other, even more important contributions, the explanation of the mystery of Saturn's rings and the discovery of his brightest satellite, Titan. In any case, the great Dutch astronomer speaks with authority of the planets. In his old age he enjoyed a philosopher's holiday and, on the pages of his *Cosmotheoros* (1), committed to posterity his enchanted visions of a universe filled with innumerable suns, attended by planets of which he could not think but as adorned by plants, animals and even beings endowed with the light of reason. Time does not permit to outline the bearing of the trend of speculative thinking characteristic of the period on this flight of imagination. To us Huygens' frequently naive speculations on the inhabitants of the planets are a mere curiosity, but his discerning remarks about the physical conditions pertaining to these celestial bodies command our respect. I would like to quote a few passages from an anonymous English translation of the *Cosmotheoros* (2), published in 1698, which was

brought to my attention by Mr. James Stokley. A modern translation is not available, to my knowledge, though a French one should be included in due time in the complete edition of Huygens' works, now being in course of publication under the auspices of the Royal Dutch Academy of Sciences.

. . . "There's no Body I suppose but sees, that in the first the Earth is made to be of the same sort with the rest of the Planets. . . . And that the other Planets are round like it, and like it receive all the Light they have from the Sun, there's no room (since the Discoveries made by Telescopes) to doubt. Another thing they are like in is, that they are moved round their own Axis; for since 'tis certain that Jupiter and Saturn are, who doubt it of the others? Again as the Earth has its Moon moving round it, so Jupiter and Saturn have theirs. Now since in so many things they thus agree, what can be more probable than that in others they agree too; and that the other Planets are as beautiful and as well stock'd with Inhabitants as the Earth? Or what shadow of a Reason can there be why they should not? . . .

. . . And, First, 'tis more than probable that the Bodies of the Planets are solid like that of our Earth, and that they don't want what we call Gravity, that Virtue, which like a Loadstone attracts whatsoever is near the Body to its Center. And that they have such a Quality, their very Figure is a proof; for their Roundness proceeds only from an equal pressure of all their Parts tending to the same Center." . . .

Huygens must be exonerated from the objectionable comparison of gravity with magnetism. This is merely an interpolation of the translator, as comparison with the original Latin text revealed. The author then goes on to discuss at length the presumable differences in the bodily organization of the terrestrial animals and of the planetary population, and makes a point of the fact that the discovery of the American continent has not enriched the collections of the naturalists with any species of excessively weird shape. Speaking of the organisms he says:

. . . 'Tis more than probable that all the difference there is between us and them, springs from the greater or less distance and influence from that Fountain of Heat and Life the Sun; which will cause a difference not so much in their Form and Shape, as in their Matter and Contexture.

And as for the Matter whereof the Plants and Animals there consist, tho' it is impossible ever to come to the knowledge of its Nature, yet this we may venture to assert (there being scarce any doubt of it) that their Growth and Nourishment proceeds from some liquid Principle. . . .

. . . But since 'tis certain that the Earth and Jupiter have their Water and Clouds, there is no reason why the other Planets should be without them. I can't say that they are exactly of the same nature with our Water; but that they should be liquid their use requires, as their beauty does that they should be clear. For this Water of ours, in Jupiter and Saturn, would be frozen up instantly by reason of the vast distance of the Sun. Every Planet therefore must have its Waters of such a temper, as to be proportion'd to its heat:

Jupiter's and Saturn's must be of such nature as not to be liable to Frost; and Venus's and Mercury's of such a nature, as not to be easily evaporated by the Sun. But in all of them, for a continual supply of Moisture, whatever Water is drawn up by the Heat of the Sun into Vapors, must necessarily return back again thither.

. . . The Air I confess may be much thicker and heavier than ours, and so, without disadvantage to its Transparency, be fitter for the volatile Animals. There may be too many sorts of Fluids ranged over one another in rows as it were. The Sea perhaps may have such a fluid lying on it, which tho' ten times lighter than Water, may be a hundred times heavier than Air; whose utmost Extent may not be so large as to cover the higher places of their Earth.

This may suffice. The insight that it is gravitation which moulds the shape of a planet appears now like a vision of the geodetical theories to come which relate the oblateness of a planet with the distribution of gravity over its surface. The suggestion of an atmosphere consisting of several non-miscible fluids, stratified according to their densities, anticipates the geochemical notion of the first separation of phases in the liquid primeval planets. As for the conjectures on the well-tempered waters, it must be remembered that Huygens was a contemporary of the author of the *Skeptical Chymist,* and in his days there was no reason to suspect that resourceful Nature might be restrained by the laws of quantum physics from providing waters of any desirable degree of volatility. Our immensely enlarged knowledge of the physics and chemistry of matter prevents us from sharing Huygens' optimistic views on the planets as abodes of life, and moreover makes it so difficult to attack the problem of planetary structure by the purely theoretical methods so successfully employed in analyzing the constitution of the fixed stars.

On the scale of cosmic temperatures the planets are cold bodies, consisting of matter in the condensed, *i.e.* liquid or solid, state, and the conditions existing at their surfaces can be readily visualized in terms of every-day experience. The extreme state of thermal agitation in the interior of the gaseous stars may defy human imagination, but it enables the theoretical astrophysicist to resort to the mathematically rather tractable laws of heat radiation and ideal gases, into which the chemical composition of the star enters only as a single parameter, namely the mean molecular mass. So one of the great pioneers in this field of research, Sir Arthur Eddington, came to speak of "such a simple thing as a star." At the lower end of the temperature scale, realized in such cold aggregations of cosmic matter as planets, comets and meteorites, nature displays an embarrassing variety of phenomena, integrating the simple into the complex; there appear molecules, crystals, colloidal

systems and finally living organisms. What Professor Russell has so aptly called the greatest problem of deductive physics,—namely, given a large quantity of matter, isolated in space, with all necessary initial conditions regarding it, to find its subsequent history—will scarcely admit of a rigorous solution beyond the point where condensation starts. Such a simple compound as water, for example, is known to exist in seven crystalline phases, each differing from the others in crystal structure, density and compressibility. It is true that under very high pressures, of the order of millions of atmospheres, the quantum structure of condensed matter breaks down and the limiting state thus attained can be described by a simple equation of state, but such pressures are at best reached only near the center of the biggest planets. In this so-called degenerate state matter becomes practically homogeneous again. In the condensed state the long-range interatomic forces, which are responsible for chemical bonds and crystalline structure, prevent the formation of a random spatial distribution in an atomic ensemble, and moreover may force it to split into several liquid and solid phases. Each of these has mechanical and thermal characteristics of its own, which are numerically known solely in the limited range of pressures and temperatures accessible to laboratory experiments. Application of these data to the conditions in the interior of the planets often necessitates a rather bold extrapolation. It is obvious therefore that the internal constitution of the planets cannot be dealt with analytically unless drastically simplifying assumptions are made. Such important problems as the internal cooling of a planet and the entailing shrinkage of its bulk have been studied extensively by theoretical methods, yet the results cannot be regarded as beyond debate. Under these circumstances we shall have to confine ourselves mainly to reviewing some general conclusions of preponderantly qualitative nature.

The planets may be grouped into two classes of distinctly differing characteristics, if we disregard the asteroids about which a word will be said later. Convenient names for these groups are terrestrial planets and giant planets. Incidentally, they are separated spatially by the group of the asteroids, but it is unknown at present whether this fact has any cosmogonic significance. Little is known about the recently discovered Pluto, but eventually it may be assigned to the terrestrial planets, on account of its mass and brightness. Its orbit is so eccentric that it intersects with that of Neptune. Lyttleton (3) has shown that it is not impossible that Neptune once had two satellites revolving in the same direction as the other members of the solar system, and that a close encounter between them occurred, with the result that Triton had

its direction of rotation reversed while the other satellite was ejected from the Neptunian system and became an independent member of the solar system, now known as Pluto. The terrestrial planets have masses not exceeding that of the Earth and mean densities higher than that of the average rock constituting the Earth's crust. The giant planets stand out by masses of a higher order of magnitude and by mean densities ranging between 50 and 25 per cent of that of rocks. All the relevant mechanical data (4) have been collected in Table 1. Rabe's radii are

TABLE 1

MECHANICAL DATA OF THE PLANETS

a equatorial radius, unit 6378.388 km
M mass, unit $5.966 \cdot 10^{27}$ g
g acceleration of gravity, unit 980.60 cm sec^{-2}
V velocity of escape, unit 11.188 km sec^{-1}
D mean density, unit 1 g cm^{-3}
C moment of inertia

	a	M	g	V	D	C/Ma^2
Moon	0.273	0.01226	0.1645	0.212	3.33	0.397
Mercury	0.403	0.034	0.2093	0.290	2.86	—
Venus	0.989	0.820	0.8383	0.910	4.86	—
Earth	1.000	1.000	1.0000	1.000	5.52	0.3345
Mars	0.538	0.081	0.3717	0.447	3.84	0.359
Jupiter	11.26	317.1	2.501	5.32	1.30	0.241
Saturn	9.45	95.02	1.064	3.17	0.69	0.235
Uranus	4.19	14.74	0.840	1.88	1.10	0.236
Neptune	3.89	17.27	1.141	2.12	1.62	0.241

the result of a comprehensive discussion of all available series of micrometrically measured diameters, corrected for the contrast error according to Kühl's theory. There is considerable uncertainty as regards the masses of Mercury and, to a lesser degree, of Venus. For Mercury the corresponding mean density is unreasonably small, so its mass might rather be determined from the observed radius and a plausible model of the density distribution in the interior. Such reversal of the procedure has indeed been proposed by Jeffreys. On the other hand the micrometrical diameter of Neptune, the smallest one in angular measure, may be affected by the darkening towards the limb of the planetary disk, produced by the strong methane absorption in its atmosphere. An increase of Rabe's value by $0''.1$, which is by no means improbable, would suffice to reduce the mean density of Neptune to that of Jupiter. The moment of inertia of a planetary configuration can be derived by Radau's approximation (*cf.* H. Jeffreys, *The Earth*) from the ellipticity, determined either by direct micrometrical measurement or from the

advance of the line of apsides of a satellite's orbit. For a homogeneous body C/Ma^2 has the maximum value, 2/5. The values given for the Moon, the Earth and Mars in Table 1 have been computed by Jeffreys and correspond to his models of these bodies, to be discussed below. The values for the four giant planets have been calculated from the dynamical ellipticities, as determined by H. Struve, P. Stroobant, N. Michalski and N. Lvoff, and the differences of the individual values from their mean, 0.24, can hardly be vouched for at present. Still, it seems that the giant planets have a strikingly similar internal constitution—their masses being far more concentrated toward the center than those of the terrestrial planets. A detailed discussion of the conditions at the surfaces of the planets and in their atmospheres is outside the scope of this paper. However, the theoretical temperatures, produced at the planets' *mean* distance on a black surface by the vertically incident solar radiation, are collected in Table 2, together with the observed maximum

TABLE 2

TEMPERATURES OF THE PLANETS

S cal cm^{-2} min^{-1}, solar constant at mean distance
T^* maximum equilibrium temperature on black planet
T observed maximum temperature of total radiation

	S	T^*	T
Mercury	13.01	631	685
Venus	3.73	464	330
Earth	1.95	392	—
Mars	0.84	316	285
Jupiter	0.072	173	135
Saturn	0.021	128	120
Uranus	0.0053	89	<90
Neptune	0.0022	72	—
Pluto	0.0013	62	—

ones, which have been evaluated from the total heat radiation of the planets by application of Stefan-Boltzmann's law. There is good reason to question the validity of this procedure, at least for the giant planets, because their atmospheres have a markedly selective absorption in the infra-red and cannot be treated like black bodies. Nevertheless, the temperatures found for Jupiter and Saturn seem to be of the right order of magnitude, just high enough to maintain the spectroscopically estimated amount per square cm of gaseous ammonia in equilibrium with the solid phase. On Uranus and Neptune practically all ammonia is frozen out.

The surface temperatures of the planets have an important bearing upon their ability to retain a gaseous atmosphere against the tendency of

the molecules to dissipate into interplanetary space on account of their thermal velocities, as was first recognized by Stoney in 1897. Critical temperatures corresponding to complete loss of the atmosphere within a certain period are given in Table 3, which is an adaptation of data con-

TABLE 3

Rate of Loss of Planetary Atmospheres

J. H. Jeans, Dynamical Theory of Gases, Cambridge, 1925

	Critical absolute temperature $T°$ K, corresponding to complete dissipation of the atmosphere within		
	10^3 years	10^6 years	10^9 years
Moon	30	23	19
Mercury	104	90	65
Venus	585	460	355
Earth	674	540	425
Mars	136	116	80
Jupiter	20,520	15,720	11,540
Saturn	6,500	5,260	4,040
Uranus	2,516	1.924	1,554
Neptune	2,700	2,066	1,696

The temperatures given refer to molecular hydrogen. To apply to other gases they should be enlarged proportionally to the molecular mass μ.

tained in Jeans' *Dynamical Theory of Gases* (Cambridge, 1925). At the very high critical temperatures pertinent to Jupiter and Saturn, hydrogen would be dissociated into free atoms, at least in the higher levels of the atmosphere where the total pressure is low. Hence, the temperatures tabulated for these planets should be divided by two, in order to apply to atomic hydrogen. It was Moulton who pointed out some thirty years ago, that by reasoning from such data as presented in Table 3 the basic difference between the constitution of the terrestrial and the giant planets can fully be explained. The masses of the giant planets were large enough to retain most of the lighter constituents of their original atmospheres, mainly hydrogen and helium, whereas the terrestrial planets lost the light material and now consist almost completely of elements not lighter than oxygen. This explanation presupposes that the primeval planets were rich in hydrogen and helium, which gases now account for the low mean densities of the giant planets. Such assumption was somewhat hypothetical at the time Moulton first set forth his ideas, but has meanwhile been substantiated by the quantitative spectrum analysis of the solar and stellar atmospheres. Since the planetary system is now believed to have originated from a close encounter of the sun with another star, and the cosmic abundance of the elements does not

appear to vary much from star to star, there is little uncertainty about the initial conditions from which the chemical history of the planets is to be deduced.

The tracing of the evolution of the Earth in its chemical aspects is the most general problem of the science of geochemistry. Only the outlines of its solution can be sketched here, following the investigations of V. M. Goldschmidt, G. Tammann, H. S. Washington, and others. Leaving the consideration of the earlier stages to the student of celestial mechanics and of the equilibrium figures of gaseous masses, the argument starts at the point where the primeval planet begins to condense. The first process clearly visualized is the separation of a gaseous atmosphere from a liquid core, formed by the coalescing droplets of refractory compounds. The atmosphere is made up of the chemically inert elements, like the noble gases, and gases of low critical temperature. The core consists mainly of the elements heavier than the oxygen atom, their oxides and other compounds. The subsequent fate of the core depends principally on the degree of reduction of the whole system, *i.e.* on the abundance ratio of the metals (silicon included) to oxygen and sulphur. If there is not enough oxygen to combine with all the metals, it will be bound first by silicon and the light metals of the first three columns of the periodic table, and the heavier metals will be left partially free, on account of their smaller affinity to oxygen. This is the case realized in the Earth. Since the mutual solubility of liquid silicates and iron—this being the representative heavy metal—is quite small at lower temperatures, the system will separate into two liquid phases, the silicates floating on top of the denser native metal. In the presence of plenty of sulphur, a third phase consisting of FeS would be formed, which is intermediate in density between silicates and metallic iron and, therefore, would occupy an intermediate position. The assumption that such a sulphide layer exists in the Earth's interior has been strongly advocated by Goldschmidt, but recent seismological data appear not to bear out his expectations. The distribution of the less abundant elements among the various phases has been studied extensively. According to the phase in which they are concentrated preferentially, Goldschmidt classifies the elements as atmophile (atmosphere), lithophile (silicate phase), chalcophile (intermediate sulphide layer), and siderophile (iron core). The next important stage during the process of the geochemical separation of the elements is the fractional crystallization of the silicate phase. Some of its products are accessible to inspection at the Earth's surface. These granitic rocks are thought to be residual products of the solidification of a gabbroid magma, the primary fractions of which were heavy

minerals, like olivine, ferrochromite and magnetite, now found in the deeper layers of the silicate mantle. The original atmosphere of the Earth was rapidly lost, and its present atmosphere, including the water of the oceans, is largely the product of the degassing of the solidifying magma, from which great amounts of steam and carbon dioxide were released. Later the carbon-dioxide was gradually split by the process of assimilation in green plants and free oxygen was accumulated in the atmosphere, a thesis to which great support has been given by Goldschmidt's investigations. The small amount of initial oxygen, necessary to maintain the life of the first green plants, probably has been the result of partial dissociation of steam in the hot primeval atmosphere, as Tammann pointed out. The equivalent amount of hydrogen, produced by the dissociation, was lost by dissipation, while the heavier oxygen was readily retained. In the giant planets the evolution of the atmospheres took an entirely different course, the great abundance of hydrogen favoring the formation of hydrides. All the less volatile hydrides, water in particular and the heavier hydrocarbons, are now frozen out, leaving ammonia and methane as the only permanent gases to be identified by their spectra.

Collateral evidence as to the constitution of the Earth's interior is afforded by the interpretation of seismological data. The analysis of travel speeds of earthquake waves and their amplitudes has definitely established the existence of discontinuities of density at the depths 2900 km and 480 km. The discontinuity at 2900 km, well known for a long time, has been identified with the boundary of the metallic core. The upper boundary of Goldschmidt's hypothetical sulphide layer should reveal itself by another marked discontinuity at a depth of somewhere between 1000 and 2000 km. Although there are indications of a sharp turn in the density-radius graph at 1200 km depth, but without a jump in the density itself, the observational evidence for the sulphide layer is quite insufficient. Therefore Bullen's (5) latest analysis of the Earth's interior (1936) assumes the density to vary continuously from 480 km to 2900 km. His treatment of the compressibility of the metallic core could now be improved, since H. Jensen (6) has just given a physical theory of the compressibility of metals in the range of pressure realized near the Earth's center. As to the nature of the recently discovered discontinuity at 480 km, Jeffreys has stressed that it would be premature to explain it by a change of composition of the material. The bulk of the silicate mantle, at depths greater than about 35 km, is widely believed to consist of dunite, a rock composed chiefly of olivine $(Mg, Fe)_2$-SiO_4. Now, fosterite Mg_2SiO_4, the main constituent of ordinary olivine,

is crystallographically isomorphous and chemically analogous to magnesium germanate Mg_2GeO_4, a compound made by Goldschmidt and found by him to exist in two different crystallographic modifications, a hexagonal one, and another cubic one of higher density. Bernal (7) has advanced the hypothesis that Mg_2SiO_4 also can exist in a cubic modification, denser than the ordinary hexagonal fosterite. If this be so, high pressure would favor the formation of the hexagonal lattice, and it appears quite conceivable that the discontinuity at 480 km depth is the result of this phase transition. The terrestrial data require the density to jump from 3.69 g cm^{-3} to 4.23 g cm^{-3}, at a pressure of 162,000 atmospheres, so that a direct experimental verification is not likely to be attainable for some time to come. Jeffreys (8) has attempted an astronomical test of Bernal's hypothesis. His argument supposes that all the terrestrial planets are made of the same materials, though the low density of the moon by itself suffices to show that the proportions are not the same. If the 480 km discontinuity results from a change of crystal form, it should occur at the same pressure in the other planets, but nothing of this sort could be expected if it represented a change of material. Since the pressure at the moon's center is only about 50,000 atmospheres on Bernal's hypothesis, the relevant pressure is not reached at all, and no problem arises. To try the alternative hypothesis, it seems reasonable to suppose that the mass ratio of the two substances in the Moon is the same as that of the terrestrial layers down to 480 km and between 480 and 2900 km. This would make the mean density of the moon at least 3.67 g cm^{-3}, whereas the observed value is 3.33 g cm^{-3}. Thus Bernal's hypothesis agrees well with the data for the moon. For Mercury and Venus it does not lead to verifiable inferences. The alternative hypothesis of a new material, allows us to represent Mars' mean density, but makes the new material far more abundant compared to dunite, representing the superficial layers, than found in the Moon on this assumption. On Bernal's hypothesis, Mars must have a small iron core. Figure 1 is reproduced from Jeffreys' paper and illustrates the density distributions of the terrestrial planets, which are to be regarded as the most probable ones at present. It should be noted that the diameters and masses, on which Jeffreys based his calculations, are slightly different from those contained in Table 1.

Although the preceding interpretation of the bulk of the terrestrial planets appears rather satisfactory, there remain to be explained many details of the surface phenomena. For instance, the transitory markings on Mercury's disk, noted by several experienced observers, are difficult to understand. Mercury's high surface temperature, sufficient to melt

lead, precludes the existence of an atmosphere in which clouds of dust could spread. The nature of the clouds in Venus' atmosphere, which obstruct the view of the solid surface, is also quite obscure. Since Venus' atmosphere is known to contain a great amount of CO_2, it must

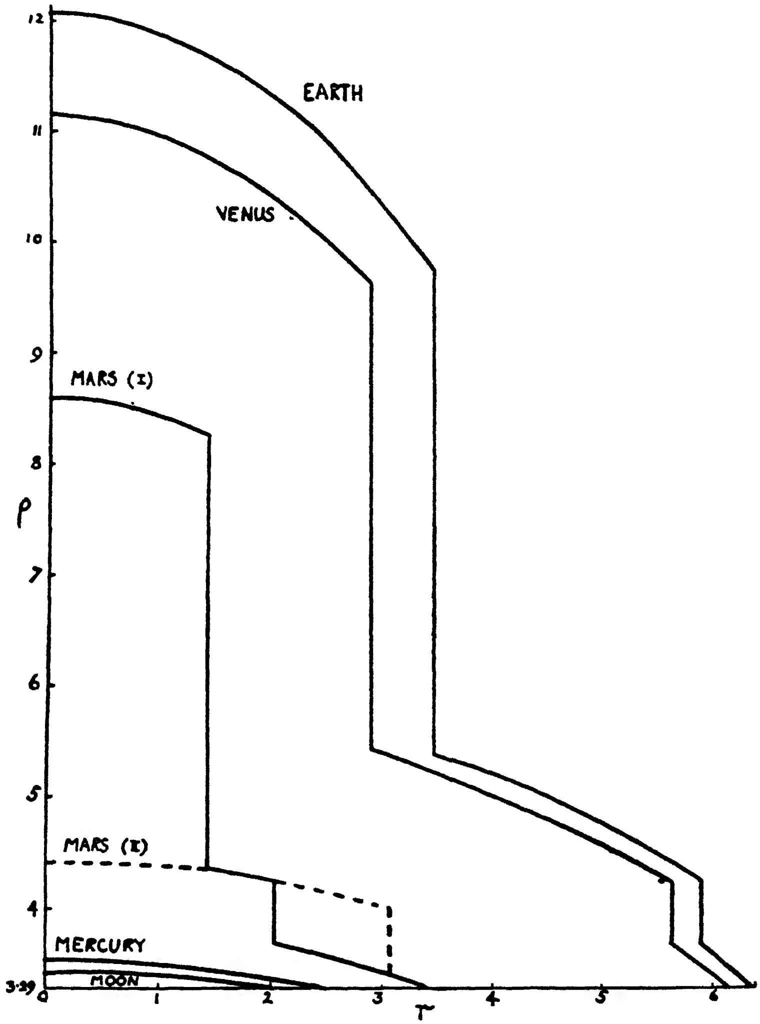


FIG. 1. Density distribution in the interior of the terrestrial planets [according to H. Jeffreys].

exert a powerful green-house effect and the temperature at the solid surface may approach the boiling point of water. Yet the spectroscopic test for water-vapor has failed to reveal as much as 1 mm of precipitable water. This shows almost conclusively that there are no oceans on

Venus. Why this planet, so similar to the Earth in every respect, has not a comparable amount of water, is one of the greatest puzzles in the chemistry of the solar system. The radiometrically determined temperatures of Mars have answered the question, whether the polar caps and clouds on this planet consist of CO_2 or H_2O, in favor of water. It is evidently much rarer than on the Earth, but this fact is easily understood because of Mars' lower critical dissipation temperature.

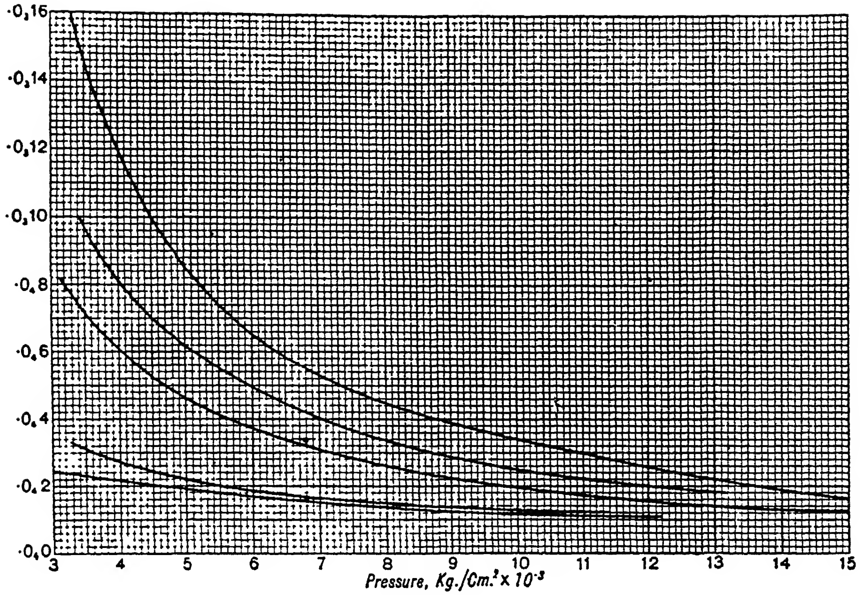
Turning now to the giant planets, the first task is to give account of low mean densities and the marked central condensation of matter in these bodies, as revealed by their moments of inertia (as noted in Table 1). By comparing the observed value of C/Ma^2 with that of a fictitious planet, supposed to have a constant density throughout, equal to the unknown surface density ρ_s, and to contain a point-mass at the center making up for the observed mean density ρ_m, Jeffreys (9) has derived the following inequality:

$$\rho_s < \frac{5}{2}\rho_m \frac{C}{Ma^2}.$$

On introducing the observed average of C/Ma^2, namely 0.24, and the mean densities from Table 1, one finds:

Jupiter: $\rho_s < 0.78$, Saturn: $\rho_s < 0.42$, Uranus: $\rho_s < 0.66$.

This at once proves that the superficial matter of the giant planets must be some extremely light substance. It is hard to believe that the giant planets could have extended atmospheres occupying a considerable part of their apparent volume. For at depths very small compared to the respective radii the atmospheres under their own weight would reach densities of the order of magnitude of those of the solidified common gases, like H_2 and He, or CH_4 and NH_3. A detailed demonstration of this fact has recently been given by Peek (10) for isothermal as well as for adiabatic atmospheres, but his numerical values represent only *lower* limits of the relevant depths, because he disregarded the decreasing compressibility of real gases, of which Fig. 2 gives a good idea. Unfortunately, the isotherms depicted in this diagram, which were measured by Bridgman (11), are the only ones known at present reaching pressures beyond the region of 3000 atmospheres, and even for these few gases there is no information available in the range of lower temperatures. In Fig. 3 there are illustrated the phase diagrams of some of the likely atmospheric constituents of the giant planets. The extrapolation of the melting curves, marked by broken lines, is based on F. Simon's (12) semi-empirical formula $\log (P + a) = b \log T + c$, which is known to represent adequately all the melting lines of numerous inorganic and



The compressibility $\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T$ as function of pressure of helium at 55° C, hydrogen at 65° C, nitrogen at 68° C, CS_2 at 65° C, and water at 65° C, reading from the top down.

FIG. 2. From P. W. Bridgman's *Physics of High Pressure.*

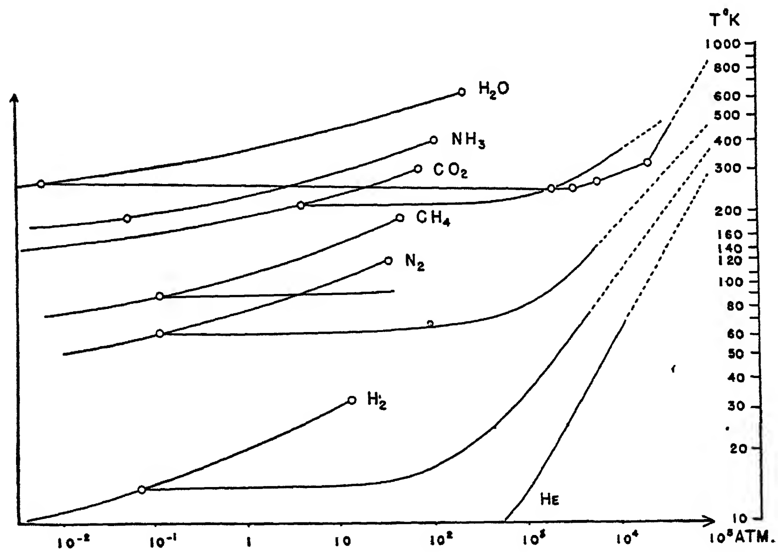


FIG. 3. Log P — Log T diagram of the phase boundaries of some common gases.

organic substances tested so far. If the assumption is justified that the atmospheres of the giant planets are isothermal, solidification of the common permanent gases under their own weight would occur at depths equivalent to pressures of about 10^4–10^5 atmospheres, as will be seen from inspection of Fig. 3. A rough estimate of the relevant depth would lead to the order of magnitude 1000 km, the exact value being quite sensitive to the unknown compressibility. But at much lower depth the fluid phase would already have lost the familiar characteristics of our atmosphere, because the compressibility approaches that of ordinary liquids (compare Fig. 2). The same may be expected of the viscosity. The general behavior of the fluid on the lower levels would probably more resemble that of the terrestrial ocean than the common notion of an atmosphere.

Proceeding on the assumption that the giant planets are isothermal bodies, their internal constitution may tentatively be represented by the following model (13): a dense core C, similar in structure to the terrestrial planets, is surrounded by a thick layer B of ice and on top of that by a layer A of highly compressed condensed gases, mainly solid hydrogen. If appropriate values for the densities of the three supposedly homogeneous layers are chosen, it is then possible to determine the two unknowns of the model, namely the radii of the surfaces of discontinuity, from the observed values of mean density and moment of inertia of the configuration. From these radii follow the volumes of the three layers, their relative masses from the assumed densities, and finally the composition of the whole configuration can be evaluated, where the core is supposed to contain 20 per cent oxygen. The summary treatment of the core, ignoring its composition of a metallic and a silicate phase, is necessary for lack of conditions other than mean density and moment of inertia as the only observables. The numerical results of the solution for Jupiter and Saturn are contained in Table 4, which also gives the corresponding pressures at the upper and lower discontinuity and at the center. The point in question is whether this model of the giant planets would put too severe demands on the abundance of hydrogen and oxygen relative to the metals compared to the relative abundance of these elements in the solar atmosphere as determined by Russell. Such is definitely not the case. An unpublished analysis of the solar atmosphere, by Menzel and Goldberg, leads to a considerable reduction of the abundance of oxygen relative to the metals. So the possibility cannot be ruled out that the primeval giant planets were not rich enough in excess oxygen to form an extended ice layer, since oxygen

would preferentially be bound by silicon and the metals on account of the higher energy of formation of these oxides.

It may be contested that the giant planets really are isothermal bodies. Most astronomers, who have devoted themselves to the study of the planetary surfaces, have found it difficult to reconcile the rapid and sometimes violent changes in the cloud formations visible on Jupiter with the thesis that the surface temperature of Jupiter is regulated solely by the amount of solar radiation received. They rather tend to assume some internal source of energy in addition. Yet Jeffreys (14) has demonstrated that Jupiter and Saturn, on every reasonable assumption about the amount of original heat energy stored in them, are old

TABLE 4

THREE-SHELL MODEL OF THE GIANT PLANETS

	Jupiter		Saturn	
	I	II	I	II
Densities (g cm^{-3}):				
A	0.35	0.25	0.35	0.25
B	1.00	1.50	1.00	1.50
C	5.50	6.00	5.50	6.00
Discontinuities:				
a, upper	0.855	0.820	0.536	0.662
b, lower	0.494	0.431	0.376	0.258
Relative Masses:				
Hydrogen	0.186	0.191	0.307	0.223
Oxygen	0.583	0.725	0.148	0.385
Metals	0.532	0.384	0.234	0.082
Pressures (atmospheres):				
P_a	$9.7 \cdot 10^5$	$9.4 \cdot 10^5$	$1.2 \cdot 10^6$	$6.6 \cdot 10^5$
P_b	$1.4 \cdot 10^7$	$1.2 \cdot 10^7$	$3.3 \cdot 10^6$	$3.8 \cdot 10^6$
P_c	$6.6 \cdot 10^7$	$5.9 \cdot 10^7$	$2.5 \cdot 10^7$	$1.6 \cdot 10^7$

enough to have cooled down by radiation to the temperature now maintained at their surfaces by the incident solar radiation. However, he does not seem to have considered the liberation of heat from radioactive material. The flux of heat from the interior at the Earth's surface is still rather uncertain, with estimates ranging from about 10^{-5}–10^{-6} cal cm^{-2} sec^{-1}. Only part of it may be of radioactive origin. An upper limit for the radioactive flux in Jupiter would result from multiplying the terrestrial value by the mass ratio, and taking account of the larger surface of Jupiter. This limiting flux, about 10^{-5} cal cm^{-2} sec^{-1}, is still small compared to the flux of solar energy at Jupiter's distance, namely $1.2 \cdot 10^{-3}$ cal cm^{-2} sec^{-1}. Nevertheless, it would take a temperature gradient of at least 1 degree per km, on the extreme assumption of

metallic conductivity, to transport this flux through the outer layers of Jupiter. Convective exchange of heat would entail an even greater temperature gradient. In any case of marked increase downwards of the temperature, it becomes somewhat doubtful whether the atmospheric temperature-pressure graph, on being inserted into Fig. 3, would intersect the line of fusion, say, of hydrogen; that means, whether there really would occur condensation of the atmosphere under its own weight. Otherwise the highly compressed fluid atmosphere might extend far into the interior until it is bounded somewhere by a new denser phase, like the silicate layer. The eventual theoretical decision between these two alternatives will depend entirely on the unknown rate of radioactive heat flux and on the mechanical and thermal properties of the atmospheric constituents in that range of high pressures which, though partially accessible to modern experimental means, is still unexplored.

There is one piece of astronomical evidence against a shallow atmosphere, viz., the drift of Jupiter's Red Spot. Relatively to the adopted zero meridian of system II the Red Spot's longitude increased from 0° in 1894 to 47° in 1903, but the direction of motion then became reversed, and, undergoing large changes of velocity, the object has passed three times through zero in the direction of diminishing longitude, viz., in 1910, 1919 and 1927. Between the oppositions of 1924 and 1925 its longitude diminished by 81° in the 13 months. Whatever the nature of this curious object may be, that it is not fixed in relation to a solid interior, is clearly shown by its drift. Since the Red Spot has now been under observation for almost 80 years, it hardly can be denied a solid nature. Any liquid would tend to spread equally all over the surface on the level corresponding to its density. The Red Spot covers an area of about 10,000 km in latitude and 45,000 km in longitude, and for a solid coherent mass its vertical extension may be expected to be of the same order of magnitude as that in latitude. If it were floating freely in the atmosphere, it could consist of some extremely light solid only. If it were dragged over solid ground, the momentum necessary to be transferred from the surrounding fluid to the huge body, in order to overcome the friction on the ground, would be considerable and might not be afforded, unless the buoyancy of the solid mass helped to reduce the friction. Obviously Jupiter's atmosphere offers fascinating mechanical and hydrodynamical problems, transcending by far in difficulty those with which the terrestrial dynamical meteorology has to struggle.

Little is known about the physical characteristics of the minor bodies of the solar system, the asteroids and meteors. In mass and size they range all the way down from the smallest satellites to the interstellar

dust. Needless to say even the largest asteroids cannot retain an atmosphere. Their main physical characteristics are apparent brightness and color index, the deviation of which from that of the Sun's might tell something about the selective reflection of these bodies. Watson and Krinov (15) have just published determinations of the reflectivity of various meteorite specimens. According to their findings, there is scarcely any chance of distinguishing between stony and metallic cosmic bodies judging by color index data, because both groups are rather uniformly grey. Two Japanese astronomers, Suzuki and Nagasima (16), claim that the frequency distribution of diameters of the asteroids, as estimated from their luminosities, agrees with that of the fractions of spherical bodies crushed by artificial impacts in laboratory experiments. The implication is that the asteroids are the debris of some early catastrophe in the solar system. On the other hand, Goldschmidt (17) has pointed out that the petrographic structure of certain meteorites gives definite proof of their solidification not having taken place in some very massive body, since the so-called pallasites consist of a nickel-iron alloy in which are found suspended droplets of silicates, indicating that the gravitational field was not strong enough to enforce separation of the liquid phases. In conclusion of this review of the constitution of planetary bodies there may be mentioned that the low densities of several satellites of the giant planets offer an intriguing problem. Whereas most of the satellites appear to consist of rock like the Moon, mean densities less than 1 gr cm^{-3} have been found for Jupiter's 4th satellite and for Mimas, Enceladus and Tethys. It is difficult to understand how small masses of such light material, presumably similar to that of the outer layers of the giant planets, can have condensed before dissipating completely like the Moon's atmosphere.

References

1. Christianii Hugenii *Cosmotheoros*, Sive De Terris Coelestibus, earumque ornatu, Conjecturae Ad Constantinum Hugenium, Fratrem: Gulielmo III. Magnae Britanniae Regi, A Secretis. Editio Altera. Hagae—Comitum, Apud Adrianum Moetjens, Bibliopolam. MDCXCIX.
2. The Celestial Worlds Discover'd: Or, Conjectures Concerning the Inhabitants, Plants and Productions of the Worlds in the Planets. Written in Latin by Christianus Huygens, And inscrib'd to his Brother Constantine Huygens, Late Secretary to his Majesty K. William. London: Printed for Timothy Childe at the White Hart at the West-end of St. Paul's Church-yard. MDCXCVIII.
3. R. A. Lyttleton, *Monthly Notices R.A.S.*, **97**, 108, 1936.
4. Radii: W. Rabe, *Astr. Nachr.*, **234**, 200, 1928. Masses: C. A. van den Bosch, Thesis, University of Utrecht, 1927.
5. K. E. Bullen, *Monthly Notices R.A.S., Geophys. Suppl.*, **3**, 395, 1936.
6. H. Jensen, *Z. f. Physik*, **111**, 373, 1938.
7. J. D. Bernal, *Observatory*, **59**, 268, 1936.
8. H. Jeffreys, *Monthly Notices R.A.S., Geophys. Suppl.*, **4**, 62, 1937.

9. H. Jeffreys, *Monthly Notices R.A.S.*, **84**, 534, 1924.
10. B. M. Peek, *Monthly Notices R.A.S.*, **97**, 574, 1937.
11. P. W. Bridgman, *Proc. Amer. Acad.*, **59**, 123, 1923. See also M. Benedict, *Journ. Amer. Chem. Soc.*, **59**, 2242, 1937.
12. F. Simon, *Trans. Faraday Soc.*, **33**, 65, 1937.
13. R. Wildt, *Astrophys. Journ.*, **87**, 508, 1938.
14. H. Jeffreys, *Monthly Notices R.A.S.*, **83**, 350, 1923.
15. F. G. Watson, *Proc. Nat. Acad. Sciences*, **24**, 532, 1938. E. L. Krinov, *Comptes Rendus de l'Academie des Sciences de l'U. R. S. S.*, **20**, 269, 1938.
16. S. Suzuki and H. Nagasima, *Sci. Papers of the Institute of Physical and Chemical Research*, **34**, 1086, 1938.
17. V. M. Goldschmidt, *Die Naturwissenschaften*, **10**, 918, 1922.

THE INTERNAL CONSTITUTION OF THE STARS

S. CHANDRASEKHAR

Yerkes Observatory, University of Chicago

(*Read February 17, 1939, in Symposium on Progress in Astrophysics*)

IN THIS paper an attempt is made to describe the general methods that have been developed to determine the physical conditions in stellar interiors. In view of the complexity of the problem, it is of value to consider, in the first instance, only those methods which involve the minimum of assumptions. Three such methods have been developed in recent years. They are

I. The method of the integral theorems.
II. The method of the homologous transformations.
III. The method of stellar envelopes.

I. THE METHOD OF THE INTEGRAL THEOREMS

In this method the fundamental assumption is made that stars are in hydrostatic equilibrium.[1] We should then have

$$\frac{dP}{dr} = -\frac{GM(r)}{r^2}\rho, \tag{1}$$

where P is the total pressure and ρ the density at a distance r from the center and $M(r)$ is the mass enclosed inside r. Since P is the total pressure we can write

$$P = p_r + p_g, \tag{2}$$

where p_g is the gas pressure and p_r the radiation pressure; *i.e.*,

$$p_r = \tfrac{1}{3}aT^4; \qquad p_g = f(\rho, T), \tag{3}$$

where a is the Stefan-Boltzmann constant, and $f(\rho, T)$ specifies the equation of state. If the assumption is made that stellar material behaves like a perfect gas then

$$p_g = \frac{k}{\mu H}\rho T, \tag{4}$$

where k is the Boltzmann constant, μ the mean molecular mass, and H the mass of the hydrogen atom.

[1] We thus exclude from our considerations rotating and variable stars.

From the geometry of the case we have in addition to (1)

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{5}$$

Using (5) we can rewrite (1) as

$$dP = -\frac{1}{4\pi}\frac{GM(r)dM(r)}{r^4}. \tag{6}$$

A direct integration of (6) yields [2]

$$P_c = \frac{1}{4\pi}\int_0^R \frac{GM(r)dM(r)}{r^4}, \tag{7}$$

where R is the radius of the star and P_c is the central pressure. From (7) we immediately infer that

$$P_c > \frac{G}{4\pi}\int_0^R \frac{M(r)dM(r)}{R^4} = \frac{1}{8\pi}\frac{GM^2}{R^4}. \tag{8}$$

Hence with no assumption except that stars are in hydrostatic equilibrium we can set a lower limit to the central pressure. Numerically (8) reduces to

$$P_c > 4.50 \times 10^8 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \text{ atmospheres}, \tag{9}$$

where $\odot$ and $R_\odot$ refer to the mass and the radius of the sun.

We can similarly obtain a lower limit to the mean pressure defined by

$$M\bar{P} = \int_0^R P dM(r). \tag{10}$$

After an integration by parts (10) reduces to

$$M\bar{P} = -\int_0^R M(r)dP, \tag{11}$$

or using (6)

$$M\bar{P} = \frac{G}{4\pi}\int_0^R \frac{M^2(r)dM(r)}{r^4}. \tag{12}$$

From (12) we obtain without difficulty [3] that

$$\bar{P} > \frac{1}{12\pi}\frac{GM^2}{R^4} \tag{13}$$

[2] Using the boundary condition that $P = 0$ at $r = R$ the radius of the star.
[3] By replacing r by R and taking it outside the integral sign.

or numerically

$$\bar{P} > 3.0 \times 10^8 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \text{ atmospheres.} \tag{14}$$

In other words, we can expect pressures of the order of 10^9 atmospheres in the stellar interiors.

We can get somewhat sharper inequalities if we supplement our assumption of hydrostatic equilibrium by another one, namely, that the mean density $\bar{\rho}(r)$ interior to r does not increase outward. This assumption implies that

$$\bar{\rho}(r) \geqslant \rho(r). \tag{15}$$

It will be noticed that we do not exclude completely the possibility of negative density gradients. We only insist that the actual density $\rho(r)$ at any point does not exceed the mean density $\bar{\rho}(r)$ interior to the point considered.

To establish inequalities for P_c etc., we consider the following expression:

$$I_{\sigma,\nu}(r) = \int_0^r \frac{GM^\sigma(r)dM(r)}{r^\nu}. \tag{16}$$

We further assume that

$$3(\sigma + 1) - \nu > 0. \tag{17}$$

By the definition of the mean density $\bar{\rho}(r)$ we have

$$\bar{\rho}(r) = M(r)/\tfrac{4}{3}\pi r^3. \tag{18}$$

From (18) we derive that

$$r^\nu = \left\{\left(\frac{3}{4\pi}\right)\frac{M(r)}{\bar{\rho}(r)}\right\}^{\nu/3}. \tag{19}$$

Substituting (19) in (16) we have

$$I_{\sigma,\nu}(r) = G\left(\frac{4\pi}{3}\right)^{\nu/3} \int_0^r \bar{\rho}^{\nu/3}(r) M^{(3\sigma-\nu)/3}(r)dM(r). \tag{20}$$

Since according to our assumption $\bar{\rho}(r)$ does not increase outward we get an upper bound for the integral on the right hand side of (20) by replacing $\bar{\rho}(r)$ by ρ_c and taking it outside the integral sign. In the same we get a lower bound by replacing $\bar{\rho}(r)$ by its value at r[4] and taking it outside the integral sign. We obtain in this way

$$\frac{3G}{3(\sigma+1)-\nu}\left(\frac{4\pi}{3}\right)^{\nu/3} M^{(3\sigma+3-\nu)/3}(r)\bar{\rho}^{\nu/3}(r) \leqslant I_{\sigma,\nu}(r)$$
$$\leqslant \frac{3G}{3(\sigma+1)-\nu}\left(\frac{4\pi}{3}\right)^{\nu/3} M^{(3\sigma+3-\nu)/3}(r)\rho_c^{\nu/3}. \tag{21}$$

[4] r here refers to the upper bound of the integral defining $I_{\sigma,\nu}(r)$.

(21) is a fundamental inequality from which several results of importance can be derived.

From (6) we derive that

$$P_c - P = \frac{1}{4\pi} I_{1,\,4}(r). \tag{22}$$

From (21) we now infer that

$$\frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}(r)\bar{\rho}^{4/3}(r) \leqslant P_c - P \leqslant \frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}(r)\rho_c^{4/3}. \tag{23}$$

If we put $r = R$ in the above inequality, we find

$$\frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}\bar{\rho}^{4/3} \leqslant P_c \leqslant \frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}\rho_c^{4/3}. \tag{24}$$

The left hand part of the above inequality can be rewritten as

$$P_c \geqslant \frac{3}{8\pi}\frac{GM^2}{R^4} \tag{25}$$

or numerically

$$P_c > 1.35 \times 10^9 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \quad \text{atmospheres.} \tag{26}$$

(26) improves the earlier inequality (9) by a factor 3.

If we put $r = R$ in (21) we obtain

$$\frac{3}{3\sigma + 3 - \nu}\,\frac{GM^{\sigma+1}}{R^\nu} \leqslant I_{\sigma,\,\nu} \leqslant \frac{3}{3\sigma + 3 - \nu}\,\frac{GM^{\sigma+1}}{r_c^\nu}, \tag{27}$$

where r_c is defined by

$$\frac{4}{3}\pi r_c^3 \rho_c = M. \tag{28}$$

From (12) we see that

$$M\bar{P} = \frac{1}{4\pi} I_{2,\,4}. \tag{29}$$

Hence from (27) we derive that

$$\frac{3}{20\pi}\frac{GM^2}{R^4} \geqslant \bar{P} \geqslant \frac{3}{20\pi}\frac{GM^2}{r_c^4}. \tag{30}$$

The right hand part of the inequality (30) reduces to

$$\bar{P} > 5.4 \times 10^8 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \quad \text{atmospheres.} \tag{31}$$

Remembering that the potential energy Ω and the mean value of gravity

$\bar{g}$ are given by

$$-\Omega = G\int_0^R \frac{M(r)dM(r)}{r} = I_{1,1}, \tag{32}$$

and

$$M\bar{g} = G\int \frac{M(r)dM(r)}{r^2} = I_{1,2}, \tag{33}$$

we have [again using (27)] that

$$\frac{3}{5}\frac{GM^2}{R} \leqslant -\Omega \leqslant \frac{3}{5}\frac{GM^2}{r_c}, \tag{34}$$

and

$$\frac{3}{4}\frac{GM}{R^2} \leqslant \bar{g} \leqslant \frac{3}{4}\frac{GM}{r_c^2}. \tag{35}$$

The inequality for Ω further enables us to set a lower limit to the mean temperatures of gaseous stars in which radiation pressure can be neglected.[5] We define the mean temperature $\bar{T}$ by

$$M\bar{T} = \int_0^R TdM(r). \tag{36}$$

If the radiation pressure can be neglected

$$P = p_g = \frac{k}{\mu H}\rho T. \tag{37}$$

Hence from (36) and (37) we have

$$M\bar{T} = \frac{\mu H}{k}\int_0^R \frac{P}{\rho}dM(r) \tag{38}$$

$$= \frac{\mu H}{k}\int_0^R PdV, \tag{38'}$$

where dV is the volume element. However, according to a well known theorem in potential theory

$$-\Omega = 3\int_0^R PdV. \tag{39}$$

Hence combining (34), (38′) and (39) we find

$$\frac{1}{5}\frac{\mu H}{k}\frac{GM}{R} \leqslant \bar{T} \leqslant \frac{1}{5}\frac{\mu H}{k}\frac{GM}{r_c}. \tag{40}$$

The left hand side of the above inequality reduces to

$$\bar{T} \geq 4.6 \times 10^6 \mu \frac{M}{\odot}\frac{R_\odot}{R} \quad \text{degrees.} \tag{41}$$

[5] As we shall see below this is justifiable for the stars of normal mass.

In other words we can expect temperatures of the order of a few million degrees in stellar interiors.

The physical content of the results (24), (27), (34), (35) and (40) is the following: We are given a certain equilibrium configuration of mass M and radius R with an arbitrary density distribution, arbitrary except for the conditions that $\bar{\rho}(r)$ does not increase outward. From the given configuration we can construct two other configurations of uniform density—one with a constant density equal to $\bar{\rho}$ and the other with a constant density equal to ρ_c (see Fig. 1). The radii of these two

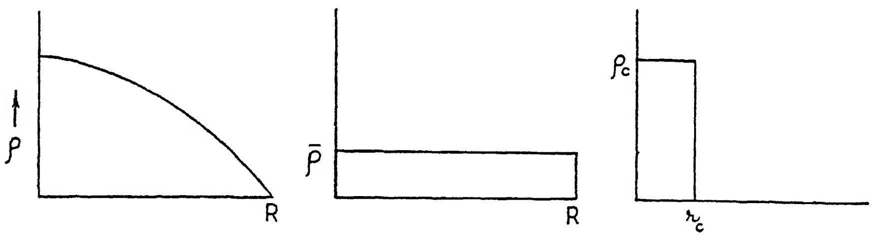


FIG. 1.

configurations are clearly R and r_c, respectively. What we have shown is that the physical variables characterizing the given equilibrium configuration, namely, P_c, $\bar{P}$, $-\Omega$, $\bar{g}$ and $\bar{T}$ (for the case of negligible radiation pressure) have values respectively less than those for the configuration of uniform density with $\rho = \rho_c$ and respectively greater than those for the configuration of uniform density with $\rho = \bar{\rho}$. Thus the given configuration is, in this sense, intermediate between the two configurations of uniform density with $\rho = \rho_c$ and $\rho = \bar{\rho}$ respectively.

Now one of the quantities which is of considerable importance in the discussion of the physical conditions in stellar interiors is the relative importance of the radiation pressure. This can be conveniently measured by the fraction $(1 - \beta)$ defined according to the relations:

$$\left.\begin{aligned} (1 - \beta)P = p_r &= \frac{1}{3} aT^4, \\ \beta P = p_g &= \frac{k}{\mu H} \rho T. \end{aligned}\right\} \tag{42}$$

By a simple elimination of T between the equations (42) we obtain

$$P = \left[\left(\frac{k}{\mu H}\right)^4 \frac{3}{a} \frac{1 - \beta}{\beta^4}\right]^{1/3} \rho^{4/3}. \tag{43}$$

Hence

$$P_c = \left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a} \frac{1 - \beta_c}{\beta_c^4}\right]^{1/3} \rho_c^{4/3}. \tag{44}$$

From (24) and (44) we now have

$$\left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a} \frac{1-\beta_c}{\beta_c^4}\right]^{1/3} \leqslant \left(\frac{\pi}{6}\right)^{1/3} G M^{2/3}, \tag{45}$$

or

$$M \geqslant \left(\frac{6}{\pi}\right)^{1/2} \left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a} \frac{1-\beta_c}{\beta_c^4}\right]^{1/2} \frac{1}{G^{3/2}}. \tag{47}$$

Define $(1 - \beta^*)$ by the equation

$$M = \left(\frac{6}{\pi}\right)^{1/2} \left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a} \frac{1-\beta^*}{\beta^{*4}}\right]^{1/2} \frac{1}{G^{3/2}}. \tag{48}$$

Combining (47) and (48) we obtain

$$\frac{1-\beta^*}{\beta^{*4}} \geqslant \frac{1-\beta_c}{\beta_c^4}. \tag{49}$$

Hence we should have

$$1 - \beta_c \leqslant 1 - \beta^*. \tag{50}$$

The inequality (50) shows that for a gaseous star the value of $(1 - \beta)$ at the center cannot exceed an amount depending on the mass of the star only. Table 1 gives the value of $(1 - \beta^*)$ for different values of the mass M. As an example of the application of Table 1, we see that for the sun, $1 - \beta_c < 0.03$ while for Capella ($M = 4.18\odot$), $1 - \beta_c < 0.2$ assuming in both cases that $\mu_c = 1$. Table 1 clearly illustrates that for the normal stars the radiation pressure as a factor in the equation of hydrostatic equilibrium can be neglected.

TABLE 1

VALUES OF $(1 - \beta^*)$

$1-\beta^*$	$\frac{M}{\odot}\mu_c^2$	$1-\beta^*$	$\frac{M}{\odot}\mu_c^2$	$1-\beta^*$	$\frac{M}{\odot}\mu_c^2$
0.01	0.56	0.14	2.77	0.35	7.67
.02	0.81	.15	2.94	.40	9.62
.03	1.01	.16	3.11	.45	12.15
.04	1.19	.17	3.28	.50	15.49
.05	1.36	.18	3.46	.55	20.06
.06	1.52	.19	3.64	.60	26.52
.07	1.68	.20	3.83	.65	36.05
.08	1.83	.21	4.02	.70	50.92
.09	1.98	.22	4.22	.75	75.89
.10	2.14	.23	4.43	.80	122.5
.11	2.29	.24	4.65	.85	224.4
.12	2.45	.25	4.87	.90	519.6
0.13	2.61	0.30	6.12	1.00	∞

A problem of importance that has not been considered so far is the question of the minimum central temperature of stars. In solving this problem we assume (1) that the density decreases outward and (2) that the temperature decreases outward. In considering this problem we regard T as a known function of P and ρ. This problem of the minimum central temperature leads to the consideration of the following more general problem.

Given that ρ and a certain function $F(P, \rho)$ both do not increase outward, what is the minimum value of $F_c \equiv F(P_c, \rho_c)$ for an equilibrium configuration of known mass M and radius R? This problem can be solved if suitable restrictions are imposed on F. We can, in fact, prove the following.[6]

Let $F(P, \rho)$ *be such that*

$$2\frac{\partial \log F}{\partial \log P} + \frac{\partial \log F}{\partial \log \rho} > 0 \qquad (P, \rho > 0), \tag{51}$$

and

$$\frac{\partial F}{\partial P} > 0. \tag{52}$$

Then in any equilibrium configuration of prescribed mass and radius in which both F *and* ρ *do not increase outward the minimum value of* F_c *is attained in the sequence of equilibrium configurations which consist of cores of constant* F *and homogeneous envelopes.*

If we choose for F the form

$$F = P\rho^{\delta}, \tag{53}$$

then the condition (51) implies that

$$\delta > -2. \tag{54}$$

Hence an immediate consequence of the theorem stated is the following: In any equilibrium configuration of prescribed mass and radius in which both ρ and $K = P/\rho^{(n+1)/n}$, $(n > 1)$, do not increase outward the minimum value of K_c is attained in the sequence of equilibrium configurations which consist of polytropic cores of index n and homogeneous envelopes.[7] It will be noticed that if we put $n = \infty$ in this special theorem we have a means of finding the lower limit to the central temperatures of stars with negligible radiation pressure. For the general problem however, we should use the theorem in its more general form. We have

$$P = \frac{k}{\mu H}\rho T + \frac{1}{3}aT^4. \tag{55}$$

[6] The details of the proof the theorem stated will be published by the author elsewhere.
[7] This special theorem has been proved by the author, see *Ap. J.*, 87, 535, 1938.

Equation (55) when regarded as implicitly determining T as a function of P and ρ specifies the function F. It is easily verified that $T(P, \rho)$ satisfies the restrictions (51) and (52). Consequently, for a star of given mass and radius we have to construct the sequence of equilibrium configurations consisting of isothermal cores and homogeneous envelopes and determine the minimum central temperature along this sequence. The theorem now asserts that this would then give the absolute minimum for T_c under the restrictions imposed. The numerical work required to determine the minimum central temperatures for stars of different masses is rather long and tedious. Only the results will be quoted:

It is found that we can write

$$T_c \geqslant \frac{1}{2} Q(M) \frac{\mu H}{k} \frac{GM}{R}, \tag{56}$$

where Q is factor which depends in a complicated way on M and μ. The following table gives the values of $Q(M)$ for certain values of M.

TABLE 2

$\frac{M}{\odot}\mu^2$	$Q(M)$
3.4	0.61
4.5	0.595
5.75	0.575
7.1	0.554
19.4	0.421

(56) can be written also as:

$$T_c \geqslant 11.5\mu Q(M) \frac{M}{\odot} \frac{R_\odot}{R} \times 10^6 \text{ degrees.} \tag{57}$$

It is found that

$$Q \to 0.64 \qquad \text{as} \qquad M \to 0, \tag{58}$$

and

$$Q \to \beta^* \qquad \text{as} \qquad M \to \infty. \tag{59}$$

From Table 2 and the inequality (57) we see that for the sun $T_c > 7.4 \times 10^6$ degrees, while for Capella $T_c > 1.8 \times 10^6$ degrees. These results again show that we can expect temperatures of at least a few million degrees in stellar interiors. Though we have found only the minimum values it is clear that the actual values must be of the same order. They may differ from the minimum values by a factor of the order 10 but we cannot expect values of an entirely different order of magnitude.

There is one other application of the theorem we are considering which is of some importance. If we put $\delta = -4/3$ in (53) then the theorem will enable us to set a *lower* limit to $(1 - \beta_c)$ for stars in which ρ

and $(1 - \beta)$ do not increase outward. The analysis of the appropriate composite configurations leads to the surprising result that the minimum value of $(1 - \beta_c)$ is the constant value which we would ascribe to a star of the same mass if it were a complete polytrope of index $n = 3$. Now a stellar model which has played a conspicuous role in the development of the subject of stellar interiors is the so-called *standard model* in which $(1 - \beta)$ is *assumed* to be a constant. The use of this model has been criticized from several directions. For this reason it is important to realize that apart from other considerations the standard model has a definite value in so far as it has a maximal (or minimal as the case may be) characteristic. We shall return later to the physical meaning of the assumption that $(1 - \beta)$ does not increase outward.

If we now assume that stars are in radiative equilibrium then the *radiative temperature gradient* is determined by

$$\frac{dp_r}{dr} = -\frac{\kappa L(r)}{4\pi c r^2}\rho, \tag{60}$$

where κ is the stellar opacity coefficient and $L(r)$ is the amount of energy in ergs per second crossing the spherical surface of radius r. Further in (60) c is the velocity of light and the other symbols have their usual meanings. From (1) and (60) we derive that

$$\frac{dp_r}{dP} = \frac{L}{4\pi cGM}\kappa\eta, \tag{61}$$

where

$$\eta = \frac{L(r)/M(r)}{L/M} = \frac{\bar{\epsilon}(r)}{\bar{\epsilon}}. \tag{62}$$

The physical meaning of η is that it is the ratio of the average rate of liberation of energy $\bar{\epsilon}(r)$ interior to the point r to the corresponding average $\bar{\epsilon}$ for the whole star.

Integrating (61) from $r = r$ to $r = R$ and using the boundary condition $p_r = 0$ at $r = R$ we have

$$p_r = \frac{L}{4\pi cGM}\overline{\kappa\eta}(r)P, \tag{63}$$

where

$$\overline{\kappa\eta}(r) = \frac{1}{P}\int_R^r \kappa\eta dP. \tag{64}$$

Equation (63) can be written alternatively in the form

$$L = \frac{4\pi cGM(1 - \beta)}{\overline{\kappa\eta}(r)}. \tag{65}$$

If we put $r = R$ in the above equation we get a fundamental relation which enables the evaluation of L in terms of an average value of $\kappa\eta$:

$$L = \frac{4\pi cGM(1 - \beta_c)}{\overline{\kappa\eta}}. \tag{66}$$

From equation (65) we conclude that the condition $(1 - \beta)$ not increasing outward (which we used earlier in the discussion) is equivalent to the assumption of $\overline{\kappa\eta}(r)$ not increasing outward. Alternatively we should have

$$\kappa\eta(r) \geqslant [\overline{\kappa\eta}]_R{}^r. \tag{67}$$

It is clear from (67) that we can actually allow a decrease of $\kappa\eta$ (within limits) as we approach the center. In actual stellar configurations η might be expected to decrease outward, but this will not be generally true of κ. For this reason it is important to realize that the minimal characteristic of the standard model was proved under restrictions which do not exclude the possibility of $\kappa\eta$ actually decreasing outward [within the limits set by (67)].

Equation (66) enables us to set an upper limit to a mean value of the stellar opacity. For if we combine (66) with the inequality (50) we obtain

$$\overline{\kappa\eta} \leqslant \frac{4\pi cGM(1 - \beta^*)}{L}. \tag{68}$$

If we assume that $\epsilon(r)$ does not increase outward then η will not increase outward and the minimum value of η is unity. Hence

$$\overline{\kappa\eta} \geqslant \bar{\kappa}, \tag{69}$$

where

$$\bar{\kappa} = \frac{1}{P_c}\int_0^{P_c} \kappa dP. \tag{70}$$

Hence we have

$$\bar{\kappa} < \frac{4\pi cGM(1 - \beta^*)}{L}. \tag{71}$$

If we apply (71) to the case of Capella ($M = 4.18\odot$, $L = 120L_\odot$) we find that

$$\bar{\kappa}_{\text{Capella}} < 100 \text{ gm}^{-1} \text{ cm}^2. \tag{72}$$

It should be noticed that our method of averaging weights the central regions of the configuration very heavily and hence the upper limit (71) is essentially an upper limit to the opacity at the center of the configuration. The physical meaning of the inequality (71) is this: If for a star of given mass M and luminosity L, $\bar{\kappa}$ should be greater than the limit set

by (71) then either the density or the rate of generation of energy ϵ or both must increase outward in some finite regions of the interior of a star.

We thus see that no special assumptions are required to establish the orders of magnitude of the physical variables in stellar interiors.[8]

II. The Method of the Homologous Transformations

We have already seen that if we restrict ourselves to the considerations of stars of masses less than five to six times the solar mass then we can neglect the radiation pressure in the equation of hydrostatic equilibrium, *i.e.*,

$$P \simeq \frac{k}{\mu H} \rho T, \tag{73}$$

or in a somewhat better approximation we can write

$$P = \frac{k}{\bar{\beta}\mu H} \rho T, \tag{74}$$

where $\bar{\beta}$ is a certain average value of β inside the star. We assume that $\bar{\beta}$ is very nearly unity. Our two equations of equilibrium are:

$$\frac{k}{\bar{\beta}\mu H} \frac{d}{dr} (\rho T) = - \frac{GM(r)}{r^2} \rho, \tag{75}$$

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{76}$$

To make the system of equations governing the equilibrium of the star complete we need an additional equation to determine the temperature gradient. If radiative equilibrium obtains then [*cf.* equation (60)]

$$\frac{d}{dr} \left(\frac{1}{3} aT^4 \right) = - \frac{\kappa L(r)}{4\pi c r^2} \rho, \tag{77}$$

where it is clear from definition that

$$L(r) = \int_0^r 4\pi r^2 \rho \epsilon dr. \tag{78}$$

In (78) ϵ is the rate of generation of energy in ergs per second per gram of the substance. An alternative form for (77) can be obtained by combining equations (61) and (65). We have

$$\frac{dp_r}{dP} = (1 - \beta) \frac{\kappa\eta}{\overline{\kappa\eta}(r)} = \frac{p_r}{P} \frac{\kappa\eta}{\overline{\kappa\eta}(r)}, \tag{79}$$

[8] The references to the literature will be found in the author's monograph *An Introduction to the Study of Stellar Structure* (Chicago, 1939).

or in a somewhat different form:

$$\frac{dT}{T} = \frac{1}{4}\frac{\kappa\eta}{\overline{\kappa\eta}(r)}\frac{dP}{P}. \tag{80}$$

Now, according to certain well known methods of argument [9] an existing temperature is stable or unstable according as it is less than or greater than the corresponding adiabatic gradient. For an enclosure containing matter and radiation the condition of adiabacy is given by

$$dQ = d(aVT^4 + c_V T) + \left(\frac{1}{3}aT^4 + \frac{k}{\mu H}\rho T\right)dV = 0, \tag{81}$$

where c_V is the specific heat of the gas at constant volume. Equation (81) can be rewritten in the form [10]

$$\frac{dT}{T} = \frac{\Gamma_2 - 1}{\Gamma_2}\frac{dP}{P}, \tag{82}$$

where

$$\Gamma_2 = 1 + \frac{(4 - 3\beta)(\gamma - 1)}{\beta^2 + 3(\gamma - 1)(1 - \beta)(4 + \beta)}, \tag{83}$$

where γ is the ratio of the specific heats ($= 5/3$) for the gas and β has the same meaning as hitherto. The condition for the stability of the radiative gradient is therefore

$$4\frac{\Gamma_2 - 1}{\Gamma_2} \geqslant \frac{\kappa\eta}{\overline{\kappa\eta}(r)}. \tag{84}$$

The radiative gradient becomes "unstable" if $\kappa\eta/\overline{\kappa\eta}(r)$ exceeds the value on the right hand side of the above inequality. The following table (Table 3) gives the values of $4(\Gamma_2 - 1)/\Gamma_2$ for different values of

TABLE 3

$1-\beta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$4(\Gamma_2 - 1)/\Gamma_2$ . .	1.6	1.304	1.177	1.108	1.065	1.039	1.022	1.010	1.004	1.000	1.000

$1 - \beta$. It follows from Table 3 that if the energy sources are concentrated towards the center then as we approach the central regions the radiative gradient will become unstable. For the case of vanishing radiation pressure, the nature of the steady state that will be set up if the radiative gradient becomes unstable has been investigated in detail

[9] See, for instance, the author's monograph on *An Introduction to the Study of Stellar Structure* (Chicago, 1939), pp. 222–227.
[10] See reference 9, pp. 55–59.

by Cowling and Biermann. These authors come to the conclusion that if

$$\frac{\kappa\eta}{\overline{\kappa\eta}(r)} \geqslant 1.6, \tag{85}$$

then the adiabatic gradient obtains:

$$p \propto T^{\gamma/(\gamma-1)}. \tag{86}$$

Now a rigorous attack on the equations of equilibrium (75), (76), (77) and (78) [11] is possible only if the dependence of ϵ and κ on ρ and T are known. Until recently we had no information concerning the nature of the dependence of ϵ on ρ and T. At the present time we have some notions on this [12] but even now our knowledge is by no means entirely satisfactory. It is for this reason that the analysis of stellar structure has depended so largely on the study of stellar models. It would appear at first sight that the uncertainty in the law of energy generation is a serious matter. We can however show that *for stars in which the radiation pressure is negligible throughout the configuration*

$$L = \text{Constant}\,\frac{1}{\kappa_0}\frac{M^{5+s}}{R^s}(\mu\bar{\beta})^{7+s}, \tag{87}$$

if the rate of generation of energy ϵ and the coefficient of opacity κ follow the laws

$$\epsilon = \epsilon_0\rho^{\alpha}T^{\nu}, \qquad \kappa = \kappa_0\rho T^{-3-s}, \tag{88}$$

where α, ν and s are arbitrary. The constant in (87) *depends on the exponents α, ν* and *s*.[13]

To prove this let us first consider stars in radiative equilibrium. Then the equations of equilibrium can be written as

$$\frac{dP}{dr} = -\frac{GM(r)}{r^2}\rho, \tag{89}$$

$$\frac{dM(r)}{dr} = 4\pi r^2\rho, \tag{90}$$

$$P = \frac{k}{\bar{\beta}\mu H}\rho T, \tag{91}$$

$$\frac{dT}{dr} = -\frac{3}{4ac}\kappa_0\rho^2T^{-6-s}\frac{\int_0^r \rho^{\alpha+1}T^{\nu}r^2dr}{r^2}. \tag{92}$$

[11] Or (86) depending upon whether (84) is satisfied or not.
[12] See H. N. Russell's article in this volume.
[13] This theorem is due to B. Strömgren.

The system of equations (89)–(92) has to be solved with the boundary conditions

$$M(r) = M, \qquad \rho = 0, \qquad T = 0 \qquad \text{at} \qquad r = R, \tag{93}$$

and

$$M(r) = 0 \qquad \text{at } r = 0. \tag{94}$$

These provide four boundary conditions and since the system of equations (89)–(92) is equivalent to a single differential equation of the fourth order, it follows that there is just exactly one solution which will satisfy the boundary conditions (93) and (94). We shall now show that from such a solution we can construct another solution such that it will describe another configuration with a different M, R, and $\bar{\beta}\mu$; we shall see, in fact, that the transformations required to go over from one set of values M, R and μ to another set M_1, R_1 and μ_1 is the successive application of three elementary homologous transformations. To show this we proceed as follows:

Let the physical variables, after a general homologous transformation has been applied be denoted by attaching a suffix "1." for a general homologous transformation we should have

$$\begin{aligned} r_1 &= y^{n_1} r, & (\bar{\beta}\mu)_1 &= y^{n_5}\bar{\beta}\mu \\ P_1 &= y^{n_2} P, & T_1 &= y^{n_6} T \\ M(r_1)_1 &= y^{n_3} M(r), & (\kappa_0\epsilon_0)_1 &= y^{n_7}(\kappa_0\epsilon_0) \\ \rho_1 &= y^{n_4}\rho, \end{aligned} \tag{95}$$

where $n_1, \cdots, n_7$ are, for the present, arbitrary constants and y is the transformation constant. The exponents $n_1, \cdots, n_7$ should satisfy certain conditions, namely, those which are necessary for the continued validity of equations (89)–(92) in the suffixed variables. Substituting (95) in (89) we find that we should have

$$y^{n_2 - n_1} = y^{n_3 + n_4 - 2n_1}, \tag{96}$$

or

$$n_2 - n_1 = n_3 + n_4 - 2n_1. \tag{97}$$

In the same way equations (90), (91) and (92) yield:

$$n_3 - n_1 = 2n_1 + n_4, \tag{98}$$

$$n_2 = n_4 + n_6 - n_5, \tag{99}$$

$$n_6 - n_1 = n_7 + (\alpha + 3)n_4 - (6 + s - \nu)n_6 + n_1. \tag{100}$$

We have thus four equations between the seven unknowns. Hence, we should be able to express any four of the n's in terms of the other three. We shall choose n_1, n_3 and n_5 as the independent quantities.

Solving for n_2, n_4, n_6 and n_7 in terms of n_1, n_3 and n_5 we find that

$$n_2 = -4n_1 + 2n_3, \tag{101}$$

$$n_4 = -3n_1 + n_3, \tag{102}$$

$$n_6 = -n_1 + n_3 + n_5, \tag{103}$$

$$n_7 = -(s - \nu - 3\alpha)n_1 + (4 + s - \nu - \alpha)n_3 + (7 + s - \nu)n_5. \tag{104}$$

If we choose $n_1 = 1$, $n_3 = 0$ and $n_5 = 0$ we have a homologous transformation in which a star of a given mass M and mean molecular mass $\mu\bar{\beta}$ is expanded or contracted. In the same way, the choice $n_1 = 0$, $n_3 = 1$ and $n_5 = 0$ corresponds to an alteration of M while R and $\mu\bar{\beta}$ are kept unchanged. Finally, the choice $n_1 = 0$, $n_3 = 0$ and $n_5 = 1$ corresponds to an alteration of $\mu\bar{\beta}$ while M and R are kept unchanged.

We have now to consider how the luminosity is changed by a homologous transformation. Since

$$L = 4\pi \int_0^R r^2 \rho \epsilon dr, \tag{105}$$

we have according to our law (88) for ϵ

$$\kappa_0 L = 4\pi\kappa_0\epsilon_0 \int_0^R r^2 \rho^{\alpha+1} T^\nu dr. \tag{106}$$

Hence, by a general homologous transformation $\kappa_0 L$ alters to $(\kappa_0 L)_1$ where

$$(\kappa_0 L)_1 = y^{n_7 + 3n_1 + (\alpha+1)n_4 + \nu n_6}(\kappa_0 L), \tag{107}$$

or by (101), (102), (103) and (104)

$$(\kappa_0 L)_1 = y^{-sn_1 + (5+s)n_3 + (7+s)n_5}(\kappa_0 L). \tag{108}$$

In other words

$$L = \text{constant}\, \frac{M^{5+s}}{\kappa_0 R^s} (\mu\bar{\beta})^{7+s}. \tag{109}$$

Another relation of importance is that which is equivalent to (104):

$$\kappa_0\epsilon_0 = \text{constant}\, R^{(3\alpha+\nu-s)} M^{(4+s-\nu-\alpha)} (\mu\bar{\beta})^{7+s-\nu}. \tag{110}$$

It is clear that the constants in (109) and (110) can depend only on the exponents α, ν and s. We have thus proved the invariance of the luminosity formula for stars in radiative equilibrium. If, however, the law of energy generation is such that it leads to a sufficiently strong concentration of the energy sources towards the center then we should reach a stage when

$$\frac{\kappa\eta}{\overline{\kappa\eta}(r)} > 4\frac{\Gamma_2 - 1}{\Gamma_2}. \tag{111}$$

In other words going inward toward the interior of a star the radiative gradient will become unstable at some definite point $r = r_i$ (say). For stars with negligible radiation pressure we have

$$\frac{\kappa\eta}{\overline{\kappa\eta}(r_i)} = 1.6 \qquad (r = r_i). \tag{112}$$

For $r < r_i$, $\kappa\eta/\overline{\kappa\eta}(r) > 1.6$. Now the right hand side of (112) is a pure number, while the quantity on the left hand side is homology invariant. Hence the fraction $q = r_i/R$ of the radius at which the instability of the radiative gradient sets in, is the same for all stars with vanishing radiation pressure. The fraction q depends only on the exponents α, ν and s which occur in the laws for ϵ and κ. Further the material interior to r_i will be in convective equilibrium and we should have

$$\frac{p}{p_i} = \left(\frac{\rho}{\rho_i}\right)^{\gamma}, \tag{113}$$

where p_i and ρ_i refer to the pressure and the density at the interface, *i.e.*, at $r_i = qR$. Equation (113) is clearly homology invariant. We have thus proved the invariance of the form of the mass-luminosity-radius relation quite generally.

The next thing that we shall have to examine is the range of variation in the constant of proportionality in the relation (109) for the possible range of stellar models. In these discussions we shall restrict ourselves to the case $s = 1/2$, *i.e.*, assume for the law of opacity the form

$$\kappa = \kappa_0 \rho T^{-3.5}, \tag{114}$$

where κ_0 will depend upon the chemical composition.

We shall consider two models: (*a*) the model ϵ = constant and (*b*) the point source model $\epsilon = 0$, $r \neq 0$. The model ϵ = constant corresponds to a uniform distribution of the energy sources while the point source model corresponds to the complete concentration of the energy sources at the center of the star. We may expect that in the actual stars an energy source distribution is realized which is intermediate to these two limiting cases.

Consider first the model ϵ = constant. Remembering that we are restricting ourselves to stars with negligible radiation pressure we can rewrite (61) in the form:

$$\frac{k}{\mu H}\frac{3}{a}\frac{d(\rho T)}{d(T^4)} = \frac{4\pi c G M}{L\kappa\eta}. \tag{115}$$

For the model under consideration $\eta = 1$. Using (88) as our law of

opacity we have

$$\frac{k}{\mu H}\frac{3}{a}\frac{d(\rho T)}{d(T^4)} = \frac{4\pi c G M}{\kappa_0 L}\frac{T^{3+s}}{\rho}. \tag{116}$$

Let

$$\rho = \frac{\mu H}{k}\frac{a}{3}T^3 y. \tag{117}$$

Then we have

$$\frac{d(yT^4)}{d(T^4)} = \frac{4\pi c G M}{\kappa_0 L}\frac{k}{\mu H}\frac{3}{a}\frac{T^s}{y}, \tag{118}$$

or after some minor transformations:

$$\frac{1}{4}T\frac{dy}{dT} = \frac{4\pi c G M}{\kappa_0 L}\frac{k}{\mu H}\frac{3}{a}\frac{T^s}{y} - y. \tag{119}$$

Introduce the new variable x defined by

$$x = \frac{4\pi c G M}{\kappa_0 L}\frac{k}{\mu H}\frac{3}{a}T^s. \tag{120}$$

(119) now takes the form

$$\frac{1}{4}sxy\frac{dy}{dx} = x - y^2. \tag{121}$$

The general solution of (121) is easily seen to be

$$y^2 = \frac{8}{8+s}x + Bx^{-8/s}, \tag{122}$$

where B is a constant of integration. From (122) we see that if $s > 0$ the second term in (122) becomes rapidly small compared to the first term as we descend into the deeper layers of the star. Hence for layers not immediately near the boundary

$$y^2 \backsimeq \frac{8}{8+s}x. \tag{123}$$

From (117), (120) and (123) we now have

$$\rho = \text{constant } T^{3+\frac{1}{2}s}; \tag{124}$$

or again

$$P = \frac{k}{\mu H}\rho T = \text{constant } T^{4+\frac{1}{2}s}. \tag{125}$$

From (124) and (125) we see that

$$P = \text{constant } \rho^{(1+[1/(3+\frac{1}{2}s)])}. \tag{126}$$

In other words, the configurations are polytropes of index $n = 3 + \frac{1}{2}s$. For the physically interesting case $s = 1/2$ so that the stars on this model are polytropes of index $n = 3.25$. The march of density and temperature in such configurations is given in Table 4 (see Figs. 2 and 3).

TABLE 4

DENSITY AND TEMPERATURE DISTRIBUTIONS FOR THE MODEL ϵ = CONSTANT

ξ	ρ/ρ_c	T/T_c
0	1.000	1.000
0.4	0.918	0.974
0.8	0.719	0.903
1.2	0.495	0.805
1.6	0.311	0.698
2.0	0.184	0.594
2.4	0.105	0.500
2.8	0.0588	0.418
3.2	0.0326	0.349
3.6	0.0179	0.290
4.0	0.00970	0.240
4.4	0.00518	0.198
4.8	0.00271	0.162
5.2	0.00137	0.131
5.6	6.57×10^{-4}	0.105
6.0	2.92×10^{-4}	0.0818
6.4	1.16×10^{-4}	0.0615
6.8	3.78×10^{-5}	0.0436
7.2	8.63×10^{-6}	0.0277
7.6	8.19×10^{-7}	0.0134
8.0	2.96×10^{-11}	0.0006
8.0189	0	0

Further from the integration appropriate for the polytrope $n = 3.25$ we find that [14]

$$\left.\begin{aligned} \rho_c &= 88.15\,\bar{\rho}, \\ T_c &= 0.968\,\frac{\mu H}{k}\frac{GM}{R}, \\ P_c &= 20.37\,\frac{GM^2}{R^4}\,. \end{aligned}\right\} \tag{127}$$

Further by a simple transformation of the luminosity formula (66) we obtain [15]

$$L = 1.43 \times 10^{25}\frac{1}{\kappa_0}\frac{M^{5.5}}{R^{0.5}}\,\mu^{7.5}, \tag{128}$$

where L, M and R are expressed in the corresponding solar units.

[14] The integration for the polytrope $n = 3.25$ is given in *Ap. J.*, 89, 116, 1939.

[15] For the details of the derivation see the author's monograph (Ref. 9, pp. 322–327).

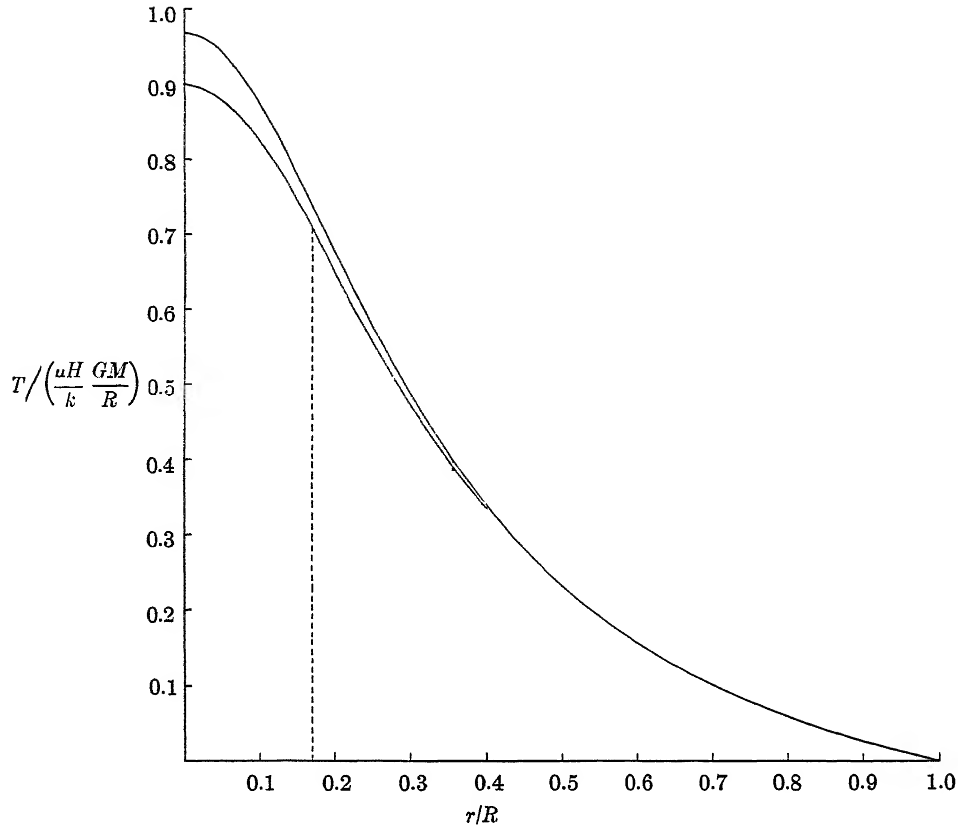


FIG. 2. The temperature distributions in the models (*a*) ϵ = constant and (*b*) the point source model.

Let us next consider the point source model. For this model it is clear that the central regions must be in convective equilibrium and therefore consist of polytropic cores of index $n = 1.5$ surrounded by "point-source envelopes," *i.e.*, regions governed by the equations

$$\frac{k}{\mu H}\frac{d}{dr}(\rho T) = -\frac{GM(r)}{r^2}\rho, \tag{129}$$

$$\frac{dp_r}{dr} = -\frac{\kappa_0 L}{4\pi c r^2}\frac{\rho^2}{T^{3.5}}, \tag{130}$$

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{131}$$

The integration has to be effected numerically. To start the integration we assume that the polytropic core extends to a fraction q of the

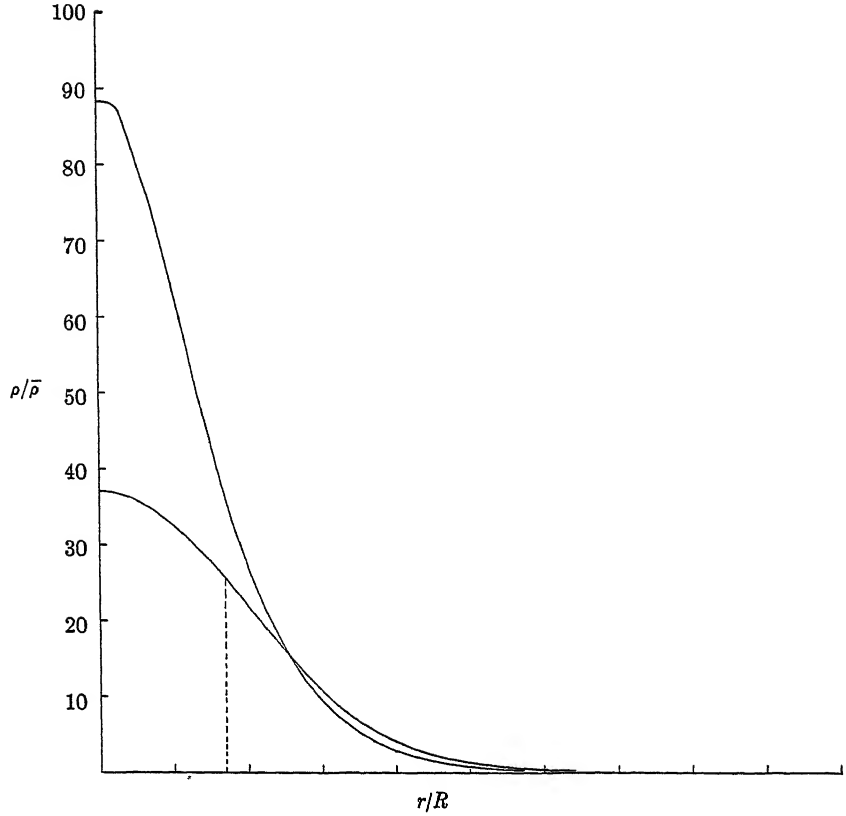


FIG. 3. The density distributions in the models (*a*) ϵ = constant and (*b*) the point source model.

radius of the star. From this point onward the integration is continued by means of equations (129)–(131). For an arbitrarily assigned fraction q, the density ρ and the temperature T will not tend to zero simultaneously. We can, however, adjust q by trial and error until ρ and T do tend to zero simultaneously. In this way we can construct a configuration of assigned mass M, radius R and mean molecular mass μ. From the homology argument it follows that the relative distributions of density, temperature, etc., will be the same for all stars with negligible radiation pressure built on this model. It is found [16] that the convective core extends to a fraction 0.17 of the radius of the star and encloses 14.5 per cent of the mass. The march of ρ and T for this model is shown in Table 5 (see Figs. 2 and 3). Further, it is found that

[16] The integration was first effected by Cowling. For greater details see the author's monograph (Ref. 9), pp. 351–355.

TABLE 5

DENSITY AND TEMPERATURE DISTRIBUTIONS FOR THE POINT-SOURCE MODEL

ξ	ρ/ρ_c	T/T_c
0	1.000	1.000
0.4	0.961	0.974
0.8	0.852	0.898
1.2	0.693	0.784
1.6	0.508	0.661
2.0	0.330	0.551
2.4	0.197	0.455
2.8	0.110	0.372
3.2	5.90×10^{-2}	0.303
3.6	3.05×10^{-2}	0.245
4.0	1.51×10^{-2}	0.196
4.4	7.16×10^{-3}	0.156
4.8	3.20×10^{-3}	0.121
5.2	1.30×10^{-3}	0.0919
5.6	4.60×10^{-4}	0.0666
6.0	1.26×10^{-4}	0.0448
6.4	2.06×10^{-5}	0.0256
6.8	6.22×10^{-7}	0.00873
7.0	5.4×10^{-10}	0.00100
7.027	0	0

$$\left.\begin{aligned} \rho_c &= 37.0\ \bar{\rho}, \\ T_c &= 0.900\,\frac{\mu H}{k}\,\frac{GM}{R}, \\ P_c &= 7.95\,\frac{GM^2}{R^4}. \end{aligned}\right\} \tag{132}$$

The luminosity formula is found to be

$$L = 5.43 \times 10^{24}\,\frac{1}{\kappa_0}\,\frac{M^{5.5}}{R^{0.5}}\,\mu^{7.5}, \tag{133}$$

where L, M and R are expressed in the corresponding solar units.

We now see that the mass-luminosity-radius relations for the two models are of the same form (in agreement with our general discussion). In addition, we notice that the constants of proportionality differ only by a factor 2.6 for the extreme range in the possible distributions of the energy sources.

For a comparison with these two models we may note that on the standard model in which stars are polytropes of index $n = 3$ we have

$$\left.\begin{aligned} \rho_c &= 54.2\ \bar{\rho}, \\ T_c &= 0.854\,\frac{\mu H}{k}\,\frac{GM}{R}, \\ P_c &= 11.05\,\frac{GM^2}{R^4}. \end{aligned}\right\} \tag{134}$$

We see that this model is "intermediate" to the two other models considered.

A fundamental result of importance which has come out of the present discussion is that there exists a relation of the type

$$L = \text{constant} \frac{1}{\kappa_0} \frac{M^{5.5}}{R^{0.5}} \mu^{7.5}, \tag{135}$$

in which the uncertainty in the constant of proportionality due to possible range of stellar models is less than the other uncertainties inherent in the problem.

As we have already pointed out κ_0 will depend on the chemical composition. The discussion of the stellar opacity is beyond the scope of the present report [17] except to mention that it is found that we can write

$$\kappa_0 = \frac{3.89 \times 10^{25} (1 - X_0^2)}{t}, \tag{136}$$

where X_0 is the hydrogen content by weight and t is a factor which is slowly varying function of ρ and T. For individual stars t can be replaced by a certain appropriate average value of $t = \bar{t}$. Thus fcr the sun it is found that $\bar{t} = 5$ while for Capella it is practically unity.

It is now obvious that an application of the relation (135) to the observational material on the masses, luminosities and the radii of the stars enables the determination of the mean molecular weight of individual stars. This problem has been investigated in great detail by Strömgren who finds a *systematic* variation of X_0 in the plane of the Hertzsprung-Russell diagram (more clearly however in a mass-radius diagram). As a result of his investigation Strömgren finds that within the limits of uncertainty of the observational material the stars can be satisfactorily arranged as a two parametric family, the two parameters being the mass M and the hydrogen content X_0. The interpretation of the Hertzsprung-Russell diagram which Strömgren arrives at is the following:

The main series up to spectral class A is the locus of stars of hydrogen content varying between 25 to 45 per cent—*i.e.*, about a mean of 35 per cent—and masses running up to 2.5 ☉. Stars of small mass and low hydrogen content are relatively rare, they occur as subgiants of spectral classes G to K. The gap between M giants and the corresponding dwarfs (on the main series) arises from the circumstance that not even stars of low hydrogen content "scatter" in this region. The massive stars ($M > 5$ ☉) occurring in the region of the B-stars which are rich in hydrogen (X_0 sometimes going up to 95 per cent)

[17] For details see the author's monograph, pp. 261–272.

form the continuation of the main series, the continuation arising from the circumstance that massive stars with " medium " hydrogen content ($0.4 < X_0 < 0.8$) which are on the main series occur in a very small region of the H.R diagram. (We shall obtain evidence in Section III for the breakdown of the model underlying these computations for the very massive stars. Further, along the main series the breakdown probably sets in at about $M = 10\odot$. The investigations of the hydrogen contents of the B-stars are therefore somewhat inconclusive.) The giant branch is characterized by stars having about the same hydrogen content as (or somewhat less than) the main series stars. The giant branch is limited on the side of low luminosity, since stars of low luminosity are relatively rare. On the side of high luminosity it is limited again, because for X_0 a little greater than 0.35 the characteristic bend of the curves of constant X_0 (see Fig. 4) disappears and also because the stars of large mass with hydrogen content greater than about 40 per cent scatter over a large area in the H.R diagram, which must, therefore, be sparsely populated. The gap in the giant branch in the region of the spectral class F is probably due to a real scarcity of stars with masses between 2.5 and $4.5\odot$. The supergiants then are interpreted as massive stars with medium hydrogen content (see Fig. 4).

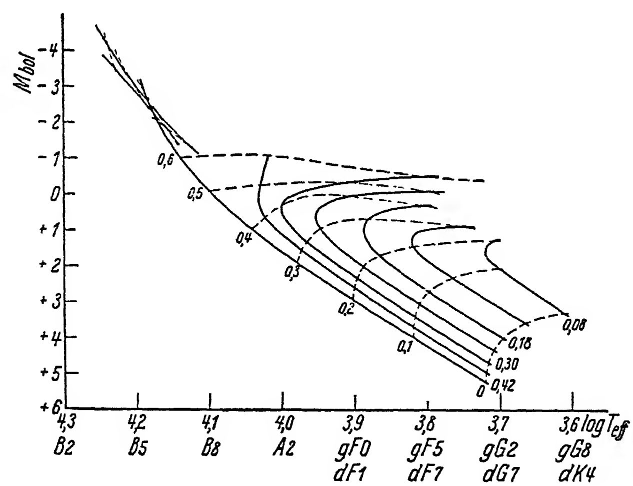


FIG. 4. Curves of constant X_0 (the full line curves) and the curves of constant M (the dotted curves) in the plane of the Hertzsprung-Russell diagram.

An important application of the method described has been recently made by Kuiper. By determining μ and X_0 for a few stars in the Hyades cluster for which he had derived the L, M and R values, he was able to

show that the stars in this cluster are relatively poor in hydrogen as compared to the normal main series stars (*i.e.*, sun, α Centauri, etc.).

So far we have considered only the relation (109). But from the homology argument we established another relation, namely (110). Now, κ_0 and ϵ_0 will depend upon the chemical composition, *i.e.*, on μ and X_0. Hence if we consider stars with a given X_0 then we can eliminate R between (109) and (110) and obtain a *pure* mass-luminosity relation. Hence for stars with constant X_0 we have:

$$L = \text{constant } M^{5+s+\frac{(4+s-\nu-\alpha)s}{3\alpha+\nu-s}}. \tag{137}$$

If we assume $s = 1/2$ and $\alpha = 2$, (137) reduces to

$$L = \text{constant } M^{\frac{63+10\nu}{11+2\nu}}. \tag{138}$$

Hence by selecting stars of a given X_0 we can determine ν by a comparison of (138) with the result of observations. This in turn will give some indications of the type of nuclear reactions that are responsible as the source of stellar energy.[18]

III. The Method of Stellar Envelopes

In this method the equilibrium of the stellar envelopes is studied. By a stellar envelope we shall mean the outer parts of a star which though containing only a small fraction (for definiteness, we shall assume this fraction to be 10 per cent) of the total mass M nevertheless occupy a good fraction of the radius R. A study of stellar envelopes has a twofold application to astrophysical theories: first, it extends the study of the conventional stellar atmospheres into the far interior and, secondly, it has also a very definite bearing on the studies of the deep interiors which are the main concern in this report. Thus the central condensation of a star, defined as the fraction ξ^* of the radius R which encloses the inner 90 per cent of the mass M, must give some indication of the concentration of mass toward the center of the star under consideration. It is clear that $(1 - \xi^*)$ is a measure of the *extent* of the stellar envelope.

In writing down the equations of equilibrium of the stellar envelope we introduce two simplifications, (*a*) that there are no sources of energy in the stellar envelope and (*b*) that the mass contained in the envelope can be neglected in comparison with the mass of the star as a whole. Indeed, these two assumptions can be regarded as defining the stellar

[18] G. Gamow, *Ap. J.*, 89, 130, 1939.

envelope. The equations of equilibrium then are

$$\frac{d}{dr}\left(\frac{k}{\mu H}\rho T + \frac{1}{3}aT^4\right) = -\frac{GM}{r^2}\rho, \tag{139}$$

$$\frac{dp_r}{dr} = -\frac{\kappa_0 L}{4\pi c r^2}\frac{\rho^2}{T^{3.5}}, \tag{140}$$

and

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{141}$$

The equations (139) and (140) can be solved explicitly [19] to give the distribution of density and temperature in the stellar envelope. Finally, equation (141) enables us to determine how far inwards we have to go to cover the first 10 per cent of the mass. The equations which determine the central condensation are found to be [20]

$$\alpha = 6.25 \times 10^{-3}\left[\frac{L^2 R\mu(1 - X_0^2)^2}{M^3}\right]^{1/4}, \tag{142}$$

$$f(\alpha; w^*) = 0.0618\frac{(1 - X_0^2)^{0.5}}{\mu^{3.75}}\left(\frac{LR^{0.5}}{M^{5.5}}\right)^{0.5}, \tag{143}$$

and

$$\xi^* = \frac{1}{(w^* + \alpha)^3\left(w^* + \frac{19}{51}\alpha\right) + 1 - \frac{19}{51}\alpha^4}. \tag{144}$$

In the above equations L, M and R are expressed in solar units. $f(\alpha; w)$ is a function defined by means of a definite integral. Tables of this function have been provided.[21]

The method of evaluating ξ^* for a star of given L, M and R and an assumed value of μ proceeds as follows:

We first determine α. Then by interpolation in the tables of the function $f(\alpha; w)$ we find the value w^* such that $f(\alpha; w^*)$ has the value given by the right hand side of (143). (144) then determines ξ^*.

Figs. 5, 6 and 7 illustrate the dependence of ξ^* on X_0 for different stars.

Without going into too much detail it is clear that the evaluation of ξ^* for the normal stars (sun, Capella, ζ Herculis) confirms the conclusions drawn by Strömgren on the basis of the standard model. To consider an example: compare the sun and ζ Herculis. Both are stars of small mass and hence of negligible radiation pressure. According to our

[19] For the solutions see the author's monograph, Chapter VIII.

[20] See reference 19.

[21] Tables of the function $f(\alpha; w)$ will be found in the author's monograph, p 361.

discussion of the homologous transformations in Section II, it is clear that these two stars must be homologous—and hence must be characterized by the same value of ξ^*. Fig. 5 shows that if ζ Herculis and the sun should have the same value, then the former must be poorer in hydrogen than the sun. This confirms the conclusion of Strömgren who has derived for the sun and ζ Herculis the values $X_0 = 0.37$ and 0.11 respectively.

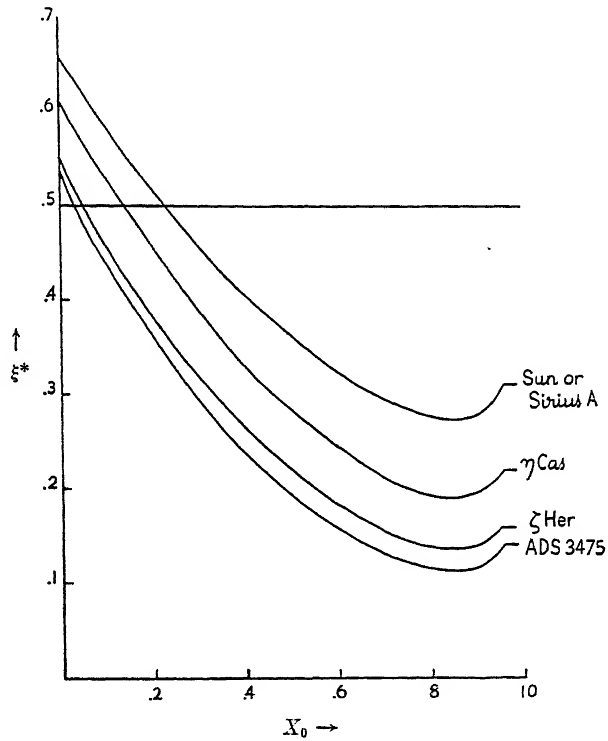


FIG. 5. (ξ^*, X_0) curves for the normal stars.

Considering now Fig. 6 we see that if we go along the sequence of stars, the sun; ζ Aurigæ B star ($M = 8.1 \odot$); μ_1 Scorpii ($M = 12 \odot$); V Puppis ($M = 18.6 \odot$); VV Cephei, B star ($M = 31 \odot$) and the Trumpler stars ($M \sim 100 \odot$) the (ξ^*, X_0) curves change continuously. This strongly suggests the breakdown of the standard model for stars on the main series sets in at about $M = 10 \odot$, becoming more and more pronounced on passing toward the larger masses. This breakdown is most clearly shown by the Trumpler stars where no adjustment of the mean molecular mass can make them homologous to the normal stars.

Turning next to Fig. 7 we notice that if we go along the sequence of stars, the sun ($M = \odot, R = R_\odot$); ζ Herculis ($M = .98 \odot$, $R = 1.9 R_\odot$);

Capella ($M = 4.2\odot$, $R = 15.8R_\odot$); ζ Aurigæ, K-star ($M = 14.8\odot$, $R = 200R_\odot$); ϵ Aurigæ, I-star ($M = 24.6\odot$, $R = 2140R_\odot$); *VV* Cephei, M-star ($M = 49\odot$, $R = 2130R_\odot$) we infer again the possibility of a breakdown of the standard model also in the region of the massive supergiants (stars of high luminosity and large radius). The

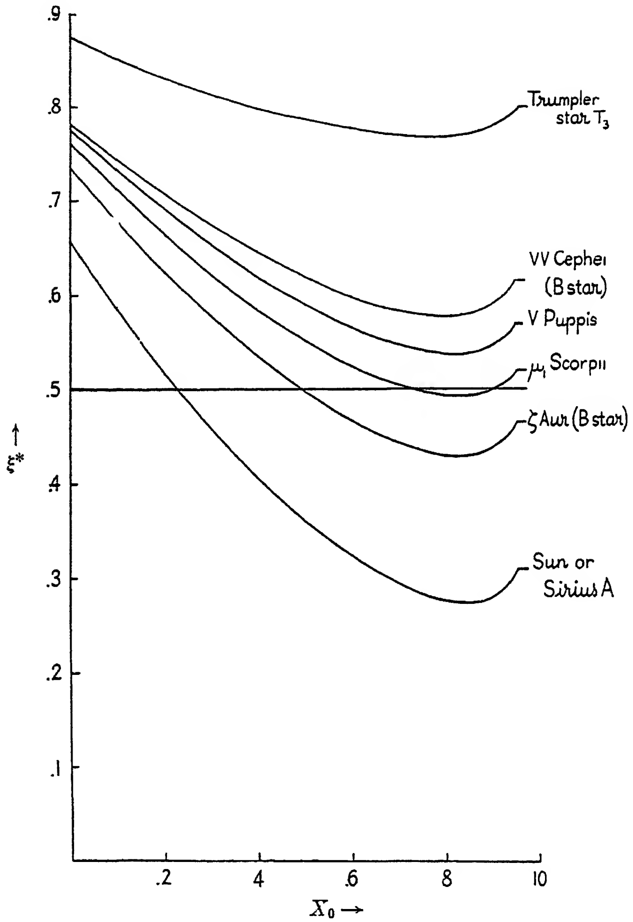


FIG. 6. (ξ^*, X_0) curves for the stars on the main series.

breakdown is now, however, in the sense of becoming more centrally condensed; this differs from the behavior of the massive stars which form an extension of the main series; the latter are certainly more homogeneous than the normal stars. Among the supergiants the possibility of finding stars with ξ^* as small as 0.05 (*e.g.*, *VV* Cephei, *M*-component) cannot be excluded.

The main results which emerge from the discussion of the central condensations of stars can be summarized as follows:

(*a*) The general way in which the theory of stellar envelopes supports the essential conclusions reached in Section II concerning the structures and the hydrogen contents of the normal stars.

(*b*) The increasing homogeneity of the massive stars on the main series, the breakdown of the standard model setting in probably at values of the mass of about 10 ⊙.

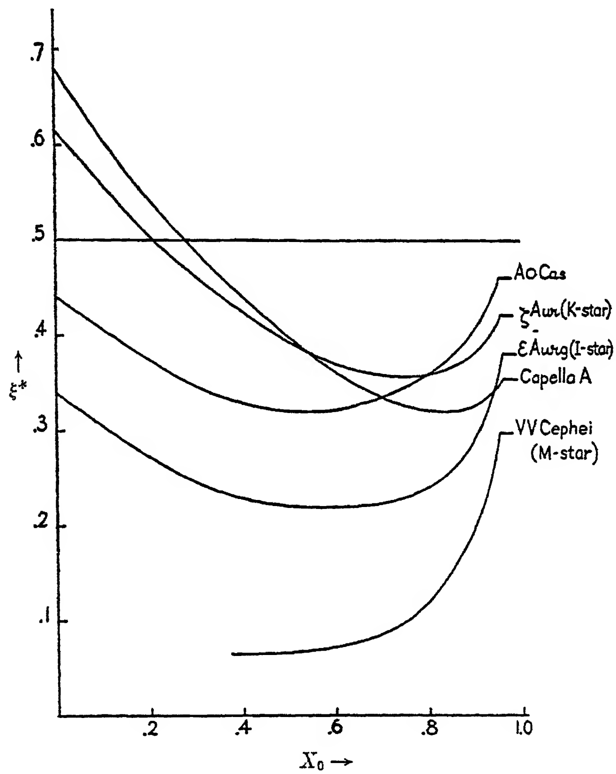


FIG. 7. (ξ^*, X_0) curves for the giants and supergiants.

(*c*) The centrally condensed nature of the massive supergiants.

We may therefore infer from the examples discussed that a certain systematic variation of the stellar model in the (M, R) plane exists.

To avoid misunderstanding, it should be pointed out that the discussion of the mass-luminosity-radius relation is satisfactory only for stars with negligible radiation pressure and consequently can be expected to be valid only for stars of normal masses ($M < 5$ ⊙). On the other hand, in the theory of stellar envelopes the effect of the radiation

pressure is taken into account "exactly" and it is for this reason that we are able to draw unambiguous conclusions concerning the structures of stars to which the standard theory cannot be applied. There is thus no contradiction involved in our discovering the "break-down" nature of the massive stars by the method of stellar envelopes.

IV. The Theory of White Dwarfs

So far we have restricted ourselves to a consideration of gaseous stars. However, as is well known, R. H. Fowler showed for the first time that the electrons assembling in the interior of the white dwarfs must be highly degenerate in the sense of the Fermi-Dirac statistics.

The theory of degeneracy in the form required for application in the theory of white dwarfs can be derived in an entirely elementary way. Now, a given number N of electrons can be confined in a given volume V by one of two methods, either by means of potential walls such that electrons inside the "potential hole" cannot escape, or by means of imposing certain periodicity conditions. We shall not consider these restrictions but the essential result of such discussions is that we can label the possible energy states for an electron inside a given volume V by means of quantum numbers in somewhat the same manner as the quantum states of an electron in an atom. If we assume that volume V is large, then it follows from the general theory that the number of quantum states with momenta between p and $p + dp$ is given by

$$V\frac{8\pi p^2 dp}{h^3}. \tag{145}$$

The meaning of (145) is simply that the accessible six-dimensional phase space can be divided into "cells" of volume h^3 and that in each cell there are two possible states. Now, the Pauli principle states that no two electrons can occupy the same quantum state. This implies that if $N(p)dp$ denotes the number of electrons with momenta between p and $p + dp$ then

$$N(p)dp \leqslant V\frac{8\pi p^2 dp}{h^3}. \tag{146}$$

A completely degenerate electron gas is one in which the lowest quantum states are all occupied. In other words, we should have

$$N(p) = V\frac{8\pi p^2}{h^3}. \tag{147}$$

It is clear that if there is only a finite number N of electrons in the specified volume, then all the electrons must have momenta less than a

certain maximum value p_0 such that

$$N = V\int_0^{p_0}\frac{8\pi p^2}{h^3}dp = \frac{8\pi V}{3h^3}p_0^3. \tag{148}$$

This "threshold value" p_0 of p is related to the electron concentration n by

$$n = \frac{N}{V} = \frac{8\pi}{3h^3}p_0^3. \tag{149}$$

To calculate the pressure, we recall that by definition the pressure P exerted by a gas is simply the mean rate of transfer of momentum across an ideal surface of unit area in the gas. From this definition it follows quite generally that

$$PV = \frac{1}{3}\int_0^\infty N(p)pv_p dp, \tag{150}$$

where v_p is the velocity associated with the momentum p. According to (147) we have for the case under consideration

$$P = \frac{8\pi}{3h^3}\int_0^{p_0} p^3\frac{\partial E}{\partial p}dp, \tag{151}$$

where E is the kinetic energy of the electron having a momentum p. According to the special theory of relativity

$$E = mc^2\left\{\left(1+\frac{p^2}{m^2c^2}\right)^{1/2} - 1\right\}, \tag{152}$$

which gives

$$\frac{\partial E}{\partial p} = \frac{1}{m}\left(1+\frac{p^2}{m^2c^2}\right)^{-1/2}p. \tag{153}$$

Substituting (153) in (151) we have

$$P = \frac{8\pi}{3mh^3}\int_0^{p_0}\frac{p^4dp}{\left(1+\frac{p^2}{m^2c^2}\right)^{1/2}}. \tag{154}$$

The integral occurring in (154) can be evaluated and we find that we can express P as

$$P = \frac{\pi m^4c^5}{3h^3}f(x), \tag{155}$$

where

$$f(x) = x(2x^2-3)(x^2+1)^{1/2} + 3\sinh^{-1}x, \tag{156}$$

and

$$x = p_0/mc. \tag{157}$$

We can now write (149) in the form

$$n = \frac{8\pi m^3c^3}{3h^3}x^3. \tag{158}$$

Equations (155), (156) and (158) represent parametrically the equation of state of a completely degenerate electron gas. (158) can be written alternatively as

$$\rho = n\mu_e H = Bx^3, \tag{159}$$

where

$$B = \frac{8\pi m^3c^3\mu_e H}{3h^3} \doteq 9.82 \times 10^5\mu_e. \tag{160}$$

Similarly we can write (155) as

$$P = Af(x), \tag{161}$$

where

$$A = \frac{\pi m^4c^5}{3h^3} = 6.01 \times 10^{22}. \tag{162}$$

Completely degenerate stellar configurations are then those which are in hydrostatic equilibrium and in which P and ρ are related according to (159) and (161). We should therefore introduce equations (159) and (161) in the equation of equilibrium,

$$\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2}{\rho}\frac{dP}{dr}\right) = -4\pi G\rho. \tag{163}$$

By the transformations

$$r = \left(\frac{2A}{\pi G}\right)^{1/2}\frac{1}{By_0}\eta; \qquad y = y_0\phi, \tag{164}$$

where

$$y_0^2 = x_0^2 + 1, \tag{165}$$

equation (163) reduces to

$$\frac{1}{\eta^2}\frac{d}{d\eta}\left(\eta^2\frac{d\phi}{d\eta}\right) = -\left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}. \tag{166}$$

Equation (166) has to be solved with the boundary conditions

$$\phi = 1; \qquad \frac{d\phi}{d\eta} = 0 \qquad \text{at} \qquad \eta = 0. \tag{167}$$

For each specified value of y_0 we have one such solution. The boundary is defined at the point where the density vanishes, and this by (159) and (165) means that if η_1 specified the boundary

$$\phi(\eta_1) = 1/y_0. \tag{168}$$

The integrations for the function ϕ have been carried out for ten different values of y_0 and the physical characteristics of the resulting configuration are shown in Table 6. The mass radius relation is shown in Fig. 8.

TABLE 6*

THE PHYSICAL CHARACTERISTICS OF COMPLETELY DEGENERATE CONFIGURATIONS

$1/y_0^2$	$M/\odot$	ρ_0 in Grams per Cubic Centimeter	ρ_{mean} in Grams per Cubic Centimeter	Radius in Centimeters
0	5.75	∞	∞	0
0.01	5.51	9.85×10^8	3.70×10^7	4.13×10^8
0.02	5.32	3.37×10^8	1.57×10^7	5.44×10^8
0.05	4.87	8.13×10^7	5.08×10^6	7.69×10^8
0.1	4.33	2.65×10^7	2.10×10^6	9.92×10^8
0.2	3.54	7.85×10^6	7.9×10^5	1.29×10^9
0.3	2.95	3.50×10^6	4.04×10^5	1.51×10^9
0.4	2.45	1.80×10^6	2.29×10^5	1.72×10^9
0.5	2.02	9.82×10^5	1.34×10^5	1.93×10^9
0.6	1.62	5.34×10^5	7.7×10^4	2.15×10^9
0.8	0.88	1.23×10^5	1.92×10^4	2.79×10^9
1.0	0	0	0	∞

* The values given in this table differ slightly from the published values (S. Chandrasekhar, *M. N.*, **95**, 208, 1935, Table III). The difference is due to the change in the accepted values of the fundamental physical constants.

The calculations are for $\mu_e = 1$. For other values of μ_e, M should be multiplied by μ_e^{-2}, R by μ_e^{-1} and ρ_0 by μ_e.

The most important characteristic of these configurations is that they possess a natural limit, *i.e.* as

$$y_0 \to \infty, \qquad \phi \to \theta_3 \tag{169}$$

(where θ_3 is the Lane-Emden function of index 3), and the mass tends to a finite limit M_3. Numerically it is found that

$$M_3 = 5.75\,\mu_e^{-2}\,\odot. \tag{170}$$

A glance at Table 6 shows that the mean density, the mass and the radius of these degenerate configurations are all of the right order of magnitude to provide the basis for the theoretical discussion of the white dwarfs. However, a really satisfactory test of the theory will consist in providing an observational basis for the existence of a mass such that as we approach it the mean density increases several times, even for a slight increase in mass. At the present time there is just one case which seems to support this aspect of the theoretical prediction. The case in question is Kuiper's white dwarf (AC 70° 8247) which is from several points of view a very remarkable star. According to Kuiper,

the most probable values of L and R are

$$\operatorname{Log} L = -1.76; \qquad \operatorname{Log} R = -2.38, \tag{171}$$

L and R being expressed in solar units. If we assume that $\mu = 1.5$, the mass-radius relation leads to a mass of 2.5 ⊙. On the other hand if we neglect relativity effects then the unrelativistic mass-radius relation (the

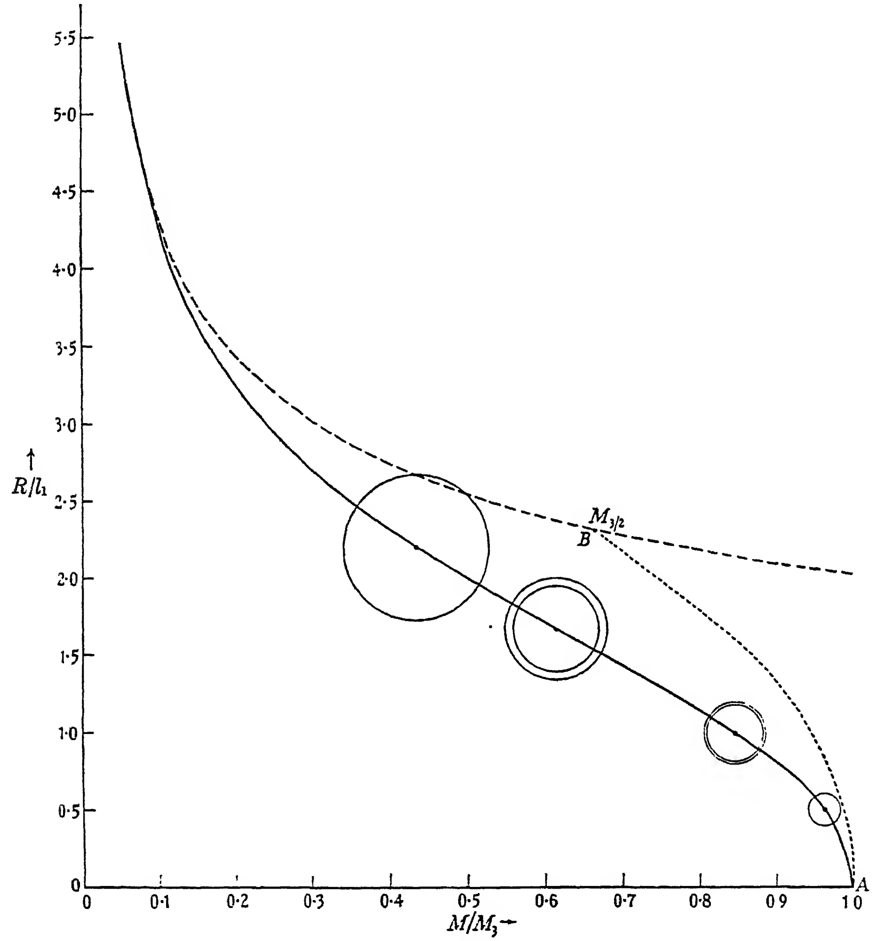


FIG. 8. The mass-radius relation for the white dwarfs.

dotted curve in Fig. 6) leads to a minimum mass of 28 ⊙. The "probable" value predicted on the unrelativistic theory will be of the order of 100 solar masses. We conclude then that the relativistic effects are confirmed by observations.[22]

[22] There is a possibility that Wolf 219, another white dwarf discovered by Kuiper, may be comparable to AC 70° 8247). If confirmed this star would be even more extraordinary than AC 70° 8247) since it is of lower luminosity.

From this brief survey of a few of the methods that have been developed in the study of stellar constitution, it should be clear that there is no reason to be sceptical of the value of such studies. Indeed, the progress that has been made in the last years, while it has been largely due to the realization of the peculiar limitations which a rational approach to the subject necessarily imposes on the investigator, justifies also the hope for still greater advances in the future.

VARIABLE STARS: A PLAN OF STUDY

CECILIA PAYNE-GAPOSCHKIN *

Phillips Astronomer, Harvard College Observatory

(Read February 17, 1939, in Symposium on Progress in Astrophysics)

The study of variable stars has become one of the important branches of modern astronomy. It offers a fertile field for the observer, and real contributions can be made to it with every type of equipment, from a pair of field glasses to the greatest of modern telescopes. In addition, it holds out almost unlimited inducements to the theoretica student of astrophysics.

The interest and flexibility of the study of variable stars have resulted in a vast accumulation of material of many kinds and qualities, and this material must be sifted before an effective use can be made of it. The critical must be separated from the uncritical, the responsible from the irresponsible. The difficulty is especially great when an observer has published only his conclusions, not the material on which they are based.

It has always been the desire of the Harvard Observatory to be alive to the development of variable star astronomy, and to obtain material that could be used in later studies of important phases of the subject. During the last few years it has seemed that the time is becoming ripe for a summing up of our knowledge of the brighter variable stars. The present contribution is intended to set forth the scope and aims of this summing up, and to show, by a few examples, the type of result that can be expected, and that will be secured, for all such stars that are bright enough to be studied.

Material and Method

A continuous photographic patrol of the whole sky has been maintained by the Harvard Observatory for almost fifty years. The present collection of about four hundred and fifty thousand photographs has been used for many years as a reference library for the study of interesting stars, but hitherto no systematic attempt has been made to extract

* The researches described in the present paper have been carried out with aid of a grant from the Milton Fund of Harvard University.

from it all the available material on variable stars, excepting for a few restricted classes of stars, such as the novæ.

The known variable stars that become as bright as, or brighter than, the tenth photographic magnitude constitute an extensive, but not prohibitive, group for systematic study. Between two and three thousand such variable stars are known at present, and there are possibly a thousand more that have been suspected of variability, but not confirmed, or that can confidently be expected to vary (for example, certain spectroscopic binaries).

Fifty-four fields in the sky, of equal area, cover the whole of the heavens, each field being comprised within one photographed region. There is a great variety in the variable star population in these fields; the sparsest field contains but four, the richest, ninety-eight stars of the kind studied. At present, work has been confined to eight or ten fields, and the results now to be described are of the nature of illustrations. It is desirable to collect all the observations and discussions of variable stars of any one type, and to postpone detailed publication until the program has been brought to completion. At that time it is important that observations be published, as well as discussions.

Details of technique would be out of place in this paper. It should be noted, however, that a uniform procedure has been adopted in the derivation of the elements of periodic variables; the phases are reckoned from J.D. 2,400,000.000, and not, as is often done, from a phase of normal maximum or minimum.

Requirements of a Scientific Attack

In planning an approach to the general problem of variable stars we have been guided by several principles. If an effective attack on any such problem is to be made, there appear to be three main requirements: the treatment must have coherence; it must utilize quantitative methods of known precision; and its results must be numerically expressed, or at least capable of numerical expression—any other results are nothing but statements of opinion.

Theoretical acceptability of results is commonly held to be an additional requirement, but it is not one that we have chosen to take into account. Not only the details, but even the general type of theoretical interpretation, is still in doubt for the intrinsic variables, and they constitute the majority of variable stars. Even for the eclipsing variables, which are comparatively well understood, there is no consensus of competent thought in the matter of the applicability of rigorous theory to the details.

The collection and coherent presentation of the facts will serve two purposes: it will direct the attention of those now studying variable stars to the salient facts; and it may tempt others to embark on theoretical investigations in an important branch of astrophysics.

The Classification of Variable Stars

In a recently published book, the writer, in collaboration,[1] has presented a scheme for the classification of variable stars. In its broader outlines the scheme is not original, but it differs in some details from most others. A transcription of it is contained in Table 1.

TABLE 1

Classification of Variable Stars

A. Geometrical Variables	*B*. The Great Sequence	*C*. Cataclysmic Variables	*D* Nebular Variables
Eclipsing stars........	Long-period variables	Novæ	
Ellipsoidal stars.......	Semiregular variables	SS Cygni stars	
Obscured stars?.......	Cepheid variables	R Coronæ Borealis stars	
	Cluster-type stars		
	(Irregular red variables)		

Obviously, a correct sense as to which characteristics of a variable star are to be considered primary, and which secondary, is essential for the formation of a significant scheme of classification. The periodicity and form of the light curve have been regarded for the most part as primary characteristics, whereas the spectrum, and hence the temperature and surface brightness, are usually treated as secondary characteristics.

It is possible that the tendency to regard the form of light change as of primary importance is a survival from the time (not very far back) when the light curves of most variable stars were almost the only data known concerning them. But to classify variable stars primarily on the basis of their spectra, though theoretically it might seem to be defensible, would nevertheless lead to confusion, by separating similar objects (such as S Cassiopeiae and R Aurigae, or RY Sagittarii and RS Telescopii), or by associating dissimilar ones (such as RU Camelopardalis and S Apodis, or R Coronae Austrinae and W Serpentis).

[1] *Variable Stars:* Harvard Observatory Monograph No. 5, 1938, by C. Payne-Gaposchkin and S. Gaposchkin.

The Eclipsing Stars—Test of Observations for Precision

The errors to which determinations of brightness from photographic plates are subject have been extensively discussed. An estimate of the precision of work of the type now to be presented has been made in the book just cited,[1] together with a brief account of the principal sources of accidental and systematic error. We need only reiterate that no magnitudes can be better than the standards on which they are based.

A practical test of the accuracy of the photographic method may be made by comparing a photographic light curve with a well-determined light curve obtained by other means. The eclipsing star δ Librae has been selected, since an excellent photoelectric light curve exists.[2] Five hundred photographic estimates, made by Miss Virginia Brenton, were used.

The observations by Stebbins are made relative to a nearby comparison star, and he publishes differences of magnitude only. For a comparison of his light curve with the photographic light curve, the constant magnitude difference was determined by means of the magnitudes of the two mean light curves at the two minima and the two maxima. The comparison is made in Table 2.

TABLE 2

Comparison of Photoelectric and Photographic Magnitudes for δ Libræ

Phase	Photoelectric	Photographic	Difference
	m	m	m
Maximum$_1$	+0.560	4.78	5.340
Minimum$_1$	−0.578	5.90	5.322
Maximum$_2$	+0.546	4.80	5.346
Minimum$_2$	+0.451	4.88	5.331
Adopted mean difference			5.335

The comparison of the photographic observations and the photoelectric measures, adjusted by means of the "adopted mean difference" of Table 2, is shown in Fig. 1. The light curve of δ Librae offers a severe test, since the star is bright ($4^m.78$ to $5^m.90$) and the estimates are therefore difficult, both because the star is about six magnitudes brighter than the limit of the plates on which it has been measured, and because

[1] C. Payne-Gaposchkin and S. Gaposchkin, *Variable Stars*, p. 349, 1938.

[2] Stebbins, *Publ. Washburn Obs.*, 15, 35, 1928. A photoelectric light curve provides a more suitable parallel than the excellent visual light curves that exist, for the effective wave-length of the photoelectric cell (about 4600 A) is nearer to the effective wave-length of the photographic system (about 4350 A) than is the effective wave-length (about 5300 A) of the visual system. The star δ Libræ suffers a partial eclipse, and the change of color at minimum is too small to affect sensibly the difference between photographic and photoelectric range.

the available comparison stars are few and not well situated. Evidently the agreement is excellent. The form and range of the photoelectric light curve are almost exactly reproduced by the photographic observations, which are competent to reveal even the slight difference of brightness between the two maxima.

Another test of the precision of the results of the photographic survey is provided by the eclipsing star RS Canum Venaticorum. Sev-

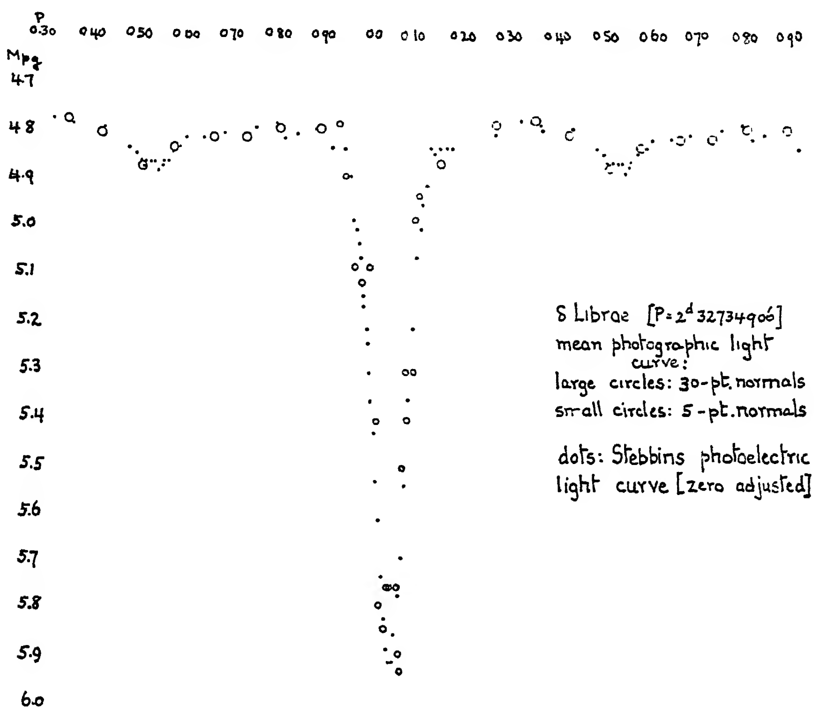


FIG. 1. Mean photographic light curve of δ Librae, compared with the photoelectric light curve published by Stebbins. Large circles represent means of thirty successive photographic observations; small circles (within the primary minimum) means of five successive observations. Dots represent individual observations by Stebbins. All observations have been reduced by means of the period $2^d.32734906$. Ordinates are magnitudes (see Table 2); abscissæ are phases, in hundredths of the period.

eral previous investigations have been made of the system, the most detailed being that published by Sitterly[1]; RS Canum Venaticorum shows a definite change of the period, and Sitterly noted a difference between the heights of the maxima, which he considered to be real. It is of interest to examine whether the photographic observations are competent to show these two features of the light curve. About four

[1] *Princeton Contr.* No. 11, 1930.

thousand photographic estimates were made for the purpose by Miss Sarah Pilsworth.

The observations during primary minimum (350 subnormal photographic observations made on Harvard photographs taken between J.D. 2,412,000 and J.D. 2,429,000) point to the change of period shown in Table 3 and in Fig. 2. When the computed phases are corrected in accordance with the changes indicated by the (empirical) smooth curve of Fig. 2, the mean light curve, as summarized in Table 4 and illustrated in Fig. 3, is derived.

TABLE 3

CHANGE OF PERIOD FOR RS CANUM VENATICORUM

Julian Date	Phase of Primary Minimum	Julian Date	Phase of Primary Minimum
	P		*P*
2420000+			
12,000–13,000	0.485	21,000–22,000	0.465
13,000–14,000	0.490:	22,000–23,000	0.460
14,000–15,000	0.488	23,000–24,000	0.460
15,000–16,000	0.492	24,000–25,000	0.465
16,000–17,000	0.500	25,000–26,000	0.460
17,000–18,000	0.490	26,000–27,000	0.470
18,000–19,000	0.475	27,000–28,000	0.472
19,000–20,000	0.475	28,000–29,000	0.472
20,000–21,000	0.470		

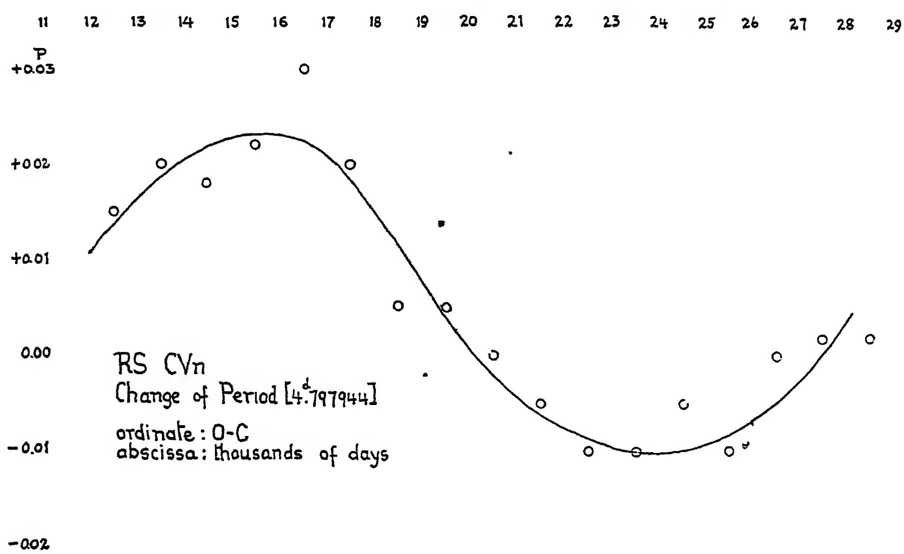


FIG. 2. Deviation of time of minimum of RS Canum Venaticorum from a uniform ephemeris (period $4^d.797944$). Ordinates are differences in phase of observed minimum from the average value $0^P.470$; abscissæ are thousands of days. The curve is arbitrarily drawn.

TABLE 4

NORMAL POINTS FOR RS CANUM VENATICORUM

Phase	Pg. Mag	No. Obs	Phase	Pg. Mag.	No Obs.	Phase	Pg. Mag.	No. Obs.
P			P			P		
0.545	8.36	100	0.225	8.34	100	0.459	9.97	10
0.572	8.36	100	0.255	8.37	100	0.469	10.02	20
0.600	8.38	100	0.280	8.40	100	0.475	10.06	10
0.630	8.36	100	0.305	8.36	100	0.480	10.03	10
0.655	8.35	100	0.345	8.34	100	0.483	10.03	10
0.680	8.34	100	0.375	8.38	100	0.488	10.04	10
0.710	8.34	100	0.400	8.38	100	0.492	9.93	10
0.740	8.34	100	0.402	8.36	10	0.495	9.85	10
0.765	8.37	100	0.406	8.41	10	0.497	9.74	10
0.790	8.34	100	0.410	8.46	10	0.501	9.65	10
0.820	8.33	100	0.414	8.46	10	0.503	9.47	10
0.850	8.35	100	0.418	8.36	10	0.505	9.51	10
0.875	8.36	100	0.422	8.46	10	0.509	8.95	10
0.905	8.36	100	0.425	8.52	10	0.514	8.90	10
0.935	8.37	100	0.429	8.57	10	0.518	8.53	10
0.970	8.39	100	0.432	8.85	10	0.522	8.47	10
0.995	8.38	100	0.435	8.79	10	0.526	8.41	10
0.020	8.34	100	0.439	9.25	10	0.533	8.42	10
0.045	8.35	100	0.441	9.37	10	0.539	8.31	8
0.065	8.33	100	0.443	9.50	10			
0.100	8.35	100	0.445	9.77	10			
0.125	8.36	100	0.447	9.84	10			
0.150	8.36	100	0.450	9.86	10			
0.180	8.36	100	0.451	9.90	10			
0.205	8.35	100	0.454	10.01	10			

The curve shown in Fig. 2 corresponds to a change of period of the same type as indicated by Fig. 1 of Sitterly's paper and by the sine term in his expression for the period; the difference between his figure and ours lies in the fact that he used the period $4^d.797875$, whereas we have used the period of $4^d.797944$, derived by Schneller.[1]

With regard to the difference in height between the two maxima, the photographic and visual observations are not in exact agreement. This is the more surprising, since the photographic observations show quite definitely the shallow secondary minimum, which is of about the same depth ($0^m.06$) as the difference observed by Sitterly between the two maxima; but the photographic observations show a difference of only about two hundredths of a magnitude between the heights of the two maxima, in the same sense as that in which it was observed by Sitterly. However, when the light curve is derived from photographic observations made at the same time as the visual observations of Sitterly, a somewhat larger difference is found. Table 5 presents the information

[1] *A.N.*, 233, 361, 1929.

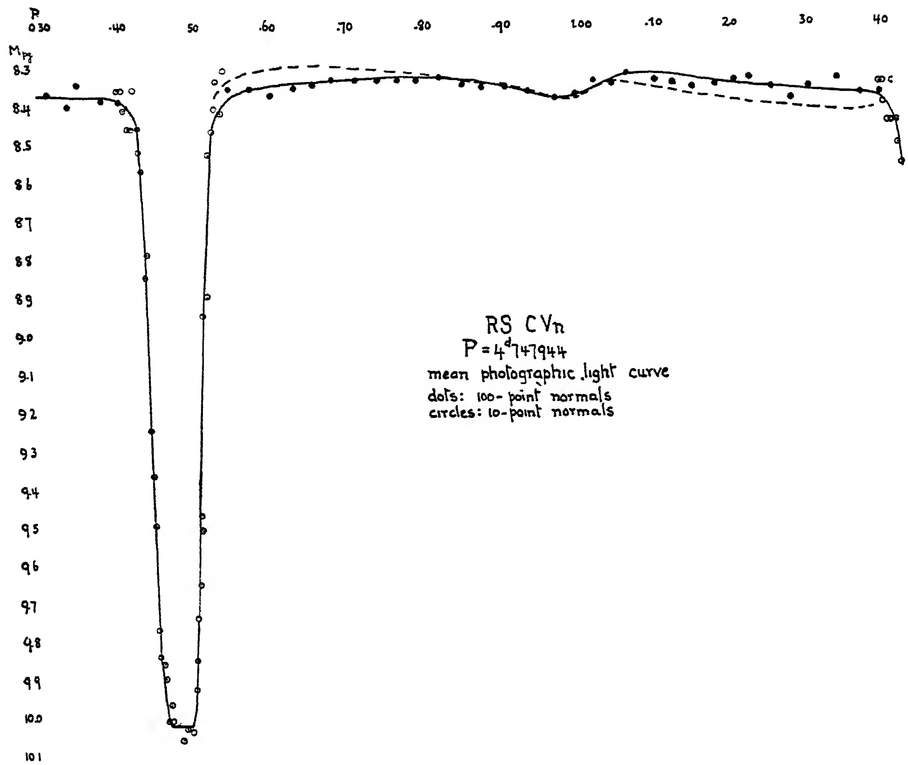


FIG. 3. Mean photographic light curve of RS Canum Venaticorum. Ordinates and abscissæ are magnitudes and phases (in terms of the period, $4^d.747944$). Dots represent means of a hundred points; circles, means of ten. The light curve is arbitrarily drawn through the observed points. The broken line between the primary minima represents the average course of Sitterly's visual light curve (zero point adjusted).

that is obtained by discussing separately the photographic observations for the times when the observed epoch of minimum (see Fig. 2) is greatest, average, and least.

It is of interest that the differences in the heights of maxima appear (from the data of Table 5) to bear some relationship to the change of

TABLE 5

RS CANUM VENATICORUM: DIFFERENCE IN HEIGHTS OF MAXIMA *

Epoch of Minimum	Photographic dM	No of Obs	Visual dM	No. of Obs.
Late (J.D. 15,000)	+0 03	340	—	—
Medium (J.D. 20,000; 28,000) . .	+0.00	910	—	—
Early (J.D. 23,000)	−0.02	750	−0.07	500?

* Maximum following primary minimum *minus* maximum following secondary.

period. The quantities are small, but considering the number of observations they still seem to be of possible significance. There is a suggestion of a presentation effect associated with orbital motion. The observed changes, if ascribed to a light-time effect, lead to an absolute orbit with a radius of about 2×10^9 kilometers; but if such an orbit is to be attributed to the action of a third body, we find that the latter should have a mass of at least 4 $\odot$, and should therefore be an observable contributor to the light of the system. Perhaps the only definite statement that can legitimately be made is that, since the difference in the heights of the maxima is apparently greater for visual than for photographic magnitudes, it should be attributed to the fainter and yellower of the two stars that constitute the system of RS Canum Venaticorum.

The eclipsing stars have been introduced principally with the purpose of serving as criteria of precision. The study of RS Canum Venaticorum has, however, suggested that the photographic observations may make significant contributions, even for well-observed stars. One further example of an eclipsing star will now be presented, in which the information obtained from the photographic survey is entirely new.

The star UW Bootis appears in the current catalogue of the elements of variable stars[1] as a W Ursae Majoris star of period $0^d.44$, with a photographic range from $9^m.0$ to $9^m.5$, and two minima of equal depth. Upon reducing the 950 photographic observations made by Miss Mary Hunt it became clear that the published elements did not represent them. A new period was accordingly derived from the dates of photographic minima. The resulting elements give a period of $1^d.0047152$, a range of $1^m.03$ at primary and of $0^m.09$ at secondary minimum. The photographic magnitude at maximum is 10.37, more than a magnitude fainter than was stated by the discoverer, and this faintness (which would have excluded the star from our systematic program, had we been aware of it) accounts for the relatively small number of observations that have been obtained.[2]

It is seen from the mean light curve (Fig. 4) that the secondary minimum is somewhat displaced from the mean position between the primary minima. A further analysis shows that the primary minimum is being continuously displaced, falling sometimes earlier, sometimes later, than the phase at which it falls in the mean light curve. The displacement of the primary minimum is corroborated by one hundred additional observations, made on several successive nights with the automatic

[1] Schneller, *Katalog und Ephemeriden Veränderlicher Sterne* für 1939.

[2] The usual number, for an individual variable that does not fall below the plate limits, is between two and four thousand.

multiple-exposure camera. Because the period is almost exactly one day, all these observations covered the same phase, which happily coincided with the primary minimum. On one side of the mean, the extreme phase of primary minimum, from these observations, is $0^{P}.73$, whereas the phase from the mean light curve is $0^{P}.77$, and the extreme phase on the other side falls at $0^{P}.82$; thus the phase of minimum has a range of nearly a tenth of the period. The relative shallowness of the primary minimum in the mean light curve, and also the apparent ellipticity of the components, are probably caused by this shifting of the primary minimum.

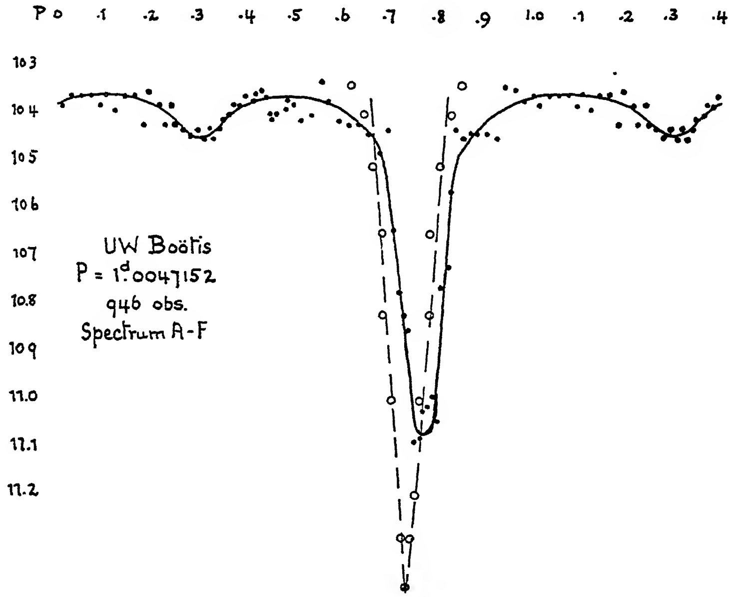


FIG. 4. Mean photographic light curve of UW Bootis. Ordinates and abscissæ are photographic magnitudes, and phases in terms of the period ($1^{d}.0047152$). Circles and a broken line represent ten-point means for a primary minimum at the extreme limit of its swing (see text).

Unfortunately the secondary minimum is too shallow for its position at any one time to be easily determined. There are indications that the period of oscillation of the primary minimum is about ten years, but the existing observations are too scanty to determine this with certainty. The period (near to one day) is an unfortunate one for a study of possible apsidal motion, since the primary and secondary minima can never be observed near the same date. The star, however, can be recommended as an interesting subject for the study of change of period, and there is little doubt that intensive visual or photographic observations would lead to interesting results.

APPLICATIONS; MEMBERS OF THE GREAT SEQUENCE

After the discussion contained in the preceding sections, which has provided a measure of precision, we can proceed with greater confidence to examine stars which have been inadequately or inconclusively studied.

An acceptable geometrical interpretation of the main facts about the eclipsing stars has long been formulated, although there are still observations that require elucidation. But for the stars of the Great Sequence there is at present no quantitative theory that fully describes the changes of brightness, spectrum, and velocity, and that also takes account of the observed relationship between the period and the luminosity. Conclusions drawn from partial data are especially dangerous when the subject is in so tentative a stage, and it is very important to marshal carefully all the available types of evidence, even for a small number of Cepheids and long-period variables. Particular attention will therefore be paid to such stars in the photographic survey. We shall here present illustrative material for three variables of the Cepheid type.

a. The Cluster Type Variable TU Ursae Majoris. Several investigators have studied this star, which is a normal cluster-type variable with a period a little greater than half a day. There are small differences between the observed or adopted periods; the value $0^d.557650$ was used by Prager,[1] $0^d.557670$ by Jacchia,[2] $0^d.55765826$ by Robinson,[3] and $0^d.557665$ by Zessewitsch,[4] Mustel,[5] Soloviev[6] and Dombrovsky.[7] A study of all available Harvard plates by Mr. Eaton has furnished 1524 estimates, the majority having been made on plates taken between J.D. 2,415,000 and J.D. 2,429,000.

The period adopted as the basis for discussion was $0^d.557665$. The observed phase of maximum, plotted against the Julian Day, is shown in Fig. 5. The period is evidently subject to fluctuations. The straight line in the figure represents the period $0^d.5576599$, which gives the best account of all the observations[8]; a small fluctuation, which might be expressed as a sine term with a period of about 12,400 days, is superimposed on the linear correction, but experience with the changes in the elements of other stars discourages us from drawing this conclusion.

[1] *Kl. Veröff.*, Ber.-Bab., No. 6, 33, 1929.
[2] *B.Z.*, No. 38, 1929.
[3] *H.A.*, 90, No. 2, 47, 1933.
[4] *B.Z.*, No. 36, 1930.
[5] *N.N.V.S.*, 4, 277, 1934.
[6] *N.N.V.S.*, 4, 383, 1935; *ibid.*, 5, 84, 1936.
[7] *Tadjik Obs. Circ.*, No. 14, 1936.
[8] Most of the other periods were deduced on the basis of observations made over only a small interval of time; Robinson's period, which was derived from about two hundred selected Harvard plates, presumably scattered throughout a large interval, is very near to the one that is here derived.

Clearly the "true" period of TU Ursae Majoris is a matter of opinion, since it will depend not only on the interval in which the observations are made, but also on the views of the investigator as to the probable nature of the fluctuation. One might, for instance, represent the times of observed maxima by three straight lines. It is our concern at present to give only a numerical summary of the observations.

b. The Cepheid Variable SU Cassiopeiae. There has been a widespread belief that the light curve of the star SU Cassiopeiae undergoes marked changes, and indeed the investigations of Robinson[1] and of

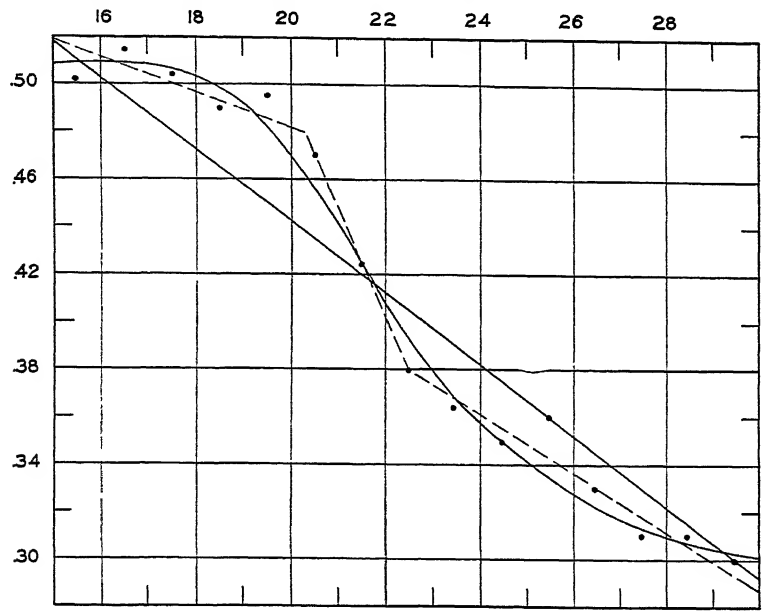


FIG. 5. Deviation of TU Ursae Majoris from a uniform ephemeris. Ordinates and abscissæ are phases, in terms of the period ($0^d.557665$), and thousands of days.

Kox[2] led respectively to ranges of $0^m.21$ and $0^m.74$. But much of this difference is to be attributed to the magnitudes adopted for the comparison stars (Table 6). When the magnitudes are reduced to a uniform system (that of the HPgP and of the HPvP, for example) the more startling differences disappear. There remains a small but definite variability in the form and amplitude of the light curve, as was also concluded by Parenago[3] from a discussion of the available visual observations.

The variations in range from all available sources, reduced to the Harvard systems of magnitudes, are summarized in Table 7. The

[1] *H.A.*, 90, No. 2, 46, 1933.
[2] *A.N.*, 256, 22, 1935.
[3] *N.N.V.S.*, 3, 7, 1930.

TABLE 6

Comparison Stars Used for SU Cassiopeiae

Photographic

H D.	B D.	HPg	Parkhurst	Robinson	Kox
		m	m	m	m
16,769	+67°224	6.15	6.15	6.59	5.88
15,784	+67 215	6.96	—	7.08	7.01
16,393	+69 171	7.50	7.45	7.60	7.52
15,849	+66 223	7.59	—	—	7.51
16,066	+67 217	7.96	—	—	8.02
16,096	+68 188	(8.05)	—	—	8.16

Visual

H.D	B.D.	HPv	Müller and Kempf	Graff	Nijland	Vogelenzang	Kox	Parenago	Zverev
		m	m	m	m	m	m	m	m
15,920	+72°140	5.07	—	—	5.31	—	—	—	—
23,277	+70 257	5.37	—	—	5.43	5.56	—	—	—
23,005	+66 284	5.73	—	—	5.88	—	—	—	—
16,024	+65 280	5.76	—	5.88	(6.43)	5.88	—	—	5.87
16,769	+67 224	5.82	6.15	6.15	6.10	6.16	5.84	6.11	6.04
15,784	+67 215	6.62	—	6.83	6.73	6.83	6.77	6.71	—
15,849	+66 223	7.17	—	—	—	—	7.27	7.19	—
16,393	+69 171	(7.37)	—	—	—	—	7.50	7.70	—
16,066	+67 217	(8.05)	—	—	—	—	7.80	—	—
16,096	+68 177	7.98	—	—	—	—	8.12	8.09	—

TABLE 7

SU Cassiopeiae: Photographic and Visual Ranges

Average J D. (thousands)	Photographic Range		Visual Range	
	m		m	
2,416,000	0.41	Gaposchkin		
17,000	0.41	Gaposchkin	0.33	Müller and Kempf
	0.44	Parkhurst	0.32	Graff
18,000	0.36	Gaposchkin	0.41	Nijland
19,000	0.35:	Gaposchkin		
20,000	0.43	Gaposchkin		
21,000	0.41	Gaposchkin	0.39	Vogelenzang
22,000	0.48:	Gaposchkin		
23,000	0.43:	Gaposchkin		
24,000	0.39	Gaposchkin		
25,000	0.44	Gaposchkin	0.32	Parenago
26,000	0.46	Gaposchkin		
27,000	0.56	Gaposchkin	0.46	Kox
	0.57	Kox		
28,000	0.48	Gaposchkin		
29,000	0.48	Gaposchkin		
Mean of all photographic:				
	0.35	Gaposchkin		
	0.34	Robinson		

Harvard plates are represented by 1600 estimates by S. Gaposchkin. There is a suggestion of periodic fluctuation of range at intervals of about 5500 days, but observations more complete and more accurate should be made to substantiate this point.[1] We should note that stars of period like that of SU Cassiopeiae, which falls in one of the minima of galactic period-frequency, are supposed to be subject to unusual irregularity in their variations of brightness.[2]

The spectral variations of SU Cassiopeiae, derived by Mrs. Margaret Mayall and the writer from 158 objective prism plates,[3] are shown in Fig. 6 and given in Table 8. The earliest spectrum occurs slightly

TABLE 8

SPECTRAL CHANGES OF SU CASSIOPEIAE

Phase *	Spectrum	No. of Obs.	Phase	Spectrum	No of Obs.
P			*P*		
0.20	F3.70	10	0.64	F3.15	10
0.23	F3.60	10	0.69	F3.60	10
0.26	F2.80	10	0.75	F4.35	10
0.29	F2.35	10	0.80	F4.15	10
0.35	F1.50	10	0.88	F4.75	10
0.41	F1.70	10	0.00	F5.30	10
0.45	F2.50	10	0.05	F6.19	8
0.52	F2.80	10	0.15	F4.20	10

* Phase of maximum is $0^{P}.50$.

before maximum brightness.[4] Designations of the color, deduced from photographic and visual light curves formed during the same interval, confirm this observation (Table 9).

The individual observations of radial velocity made at Mount Wilson [5] are plotted at the bottom of Fig. 6. On the basis of the light and color curves we should expect (from purely geometrical considerations) that the maximum radius would occur at phase $0^{P}.15$, the minimum radius at phase $0^{P}.35$. Hence we should expect a maximu velocity of recession at about phase $0^{P}.20$, where there are no ob tions of radial velocity. A redetermination of the velocity curve, a com-

[1] Such a phenomenon has been shown, for example, by the long-period variab Pavonis (Campbell and Payne, *H.B.* 872, 1930). Perhaps it is not uncommon.

[2] O'Connell, *Riverview Obs. Circ.*, 1, 1935; 2, 1936; Payne-Gaposchkin and Gaposchkin, *Variable Stars*, p. 173, 1938.

[3] These plates were taken by Mr. Daniel Norman in connection with his current spectrophotometric studies of Cepheid variables, and were kindly made available to the writer.

[4] Shapley (*Ap. J.*, 44, 278, 1916) found the earliest spectrum just after maximum light He had, however, no spectral observations just before the maximum. His spectral observations, combined into means, are plotted by dots in Fig. 5.

[5] Adams and Shapley, *Ap. J.*, 47, 46, 1918.

TABLE 9

CHANGE OF COLOR OF SU CASSIOPEIAE (J.D. 2,427,000)

Phase	Pg. Mag [1]	Pv Mag [2]	Color Index	Equivalent Spectrum [3]	Phase	Pg. Mag.	Pv. Mag.	Color Index	Equivalent Spectrum
P			m		P			m	
0.00	6.83	6.32	+0.51	F2.5	0.50	6.38	5.91	+0.47	F1.2
0.05	6.86	6.28	0.58	F4.6	0.55	6.42	5.95	0.47	F1.2
0.10	6.83	6.24	0.59	F5.0	0.60	6.50	6.00	0.50	F2.2
0.15	6.74	6.17	0.57	F4.4	0.65	6.58	6.06	0.52	F2.8
0.20	6.65	6.12	0.53	F3.1	0.70	6.64	6.11	0.53	F3.1
0.25	6.54	6.07	0.47	F1.2	0.75	6.66	6.14	0.52	F2.8
0.30	6.46	6.01	0.45	F0.5	0.80	6.67	6.16	0.51	F2.5
0.35	6.39	5.95	0.44	F0.3	0.85	6.68	6.18	0.50	F2.2
0.40	6.35	5.92	0.43	F0.0	0.90	6.73	6.22	0.51	F2.5
0.45	6.35	5.90	0.45	F0.5	0.95	6.79	6.29	0.50	F2.2

[1] Observations of Gaposchkin.
[2] Observations of Kox, reduced to Harvard system.
[3] Adopting an unpublished temperature scale for supergiant stars.

parison with concurrent observations of brightness, color, and spectrum are greatly to be desired.

c. The Cepheid Variable AL Virginis. Although bright enough to be included in Joy's study of the velocity curves of Cepheids, AL Virginis has hitherto been without a published light curve. A mean light curve, deduced from 1025 observations made by Miss Virginia Brenton, is reproduced in Fig. 7.

In the course of the reductions it became clear that the light curve was subject to fluctuations. The residuals (in magnitudes) from the mean light curve were deduced, separately, for the rising and falling branches of the light curve given in Fig. 8. These two sets of residuals, when plotted against the Julian Day, show deviations in opposite directions, thus pointing to a motion of the whole light curve parallel to the axis of abscissæ. A motion parallel to the axis of ordinates would have led to curves that were direct, and not mirror, images. The deviations in brightness, when converted (by means of the slopes of the light curve on the rising and falling branches) into fractions of the period, lead to the curve representing the deviation from a uniform ephemeris, which is shown at the bottom of Fig. 8. There can be no doubt that the period is subject to fluctuation, but it would not be possible to regard the changes as periodic.

The individual observations of radial velocity published by Joy, which are shown in the lower part of Fig. 7, have been corrected empirically by the amounts required by the curve for deviation from a uniform ephemeris.

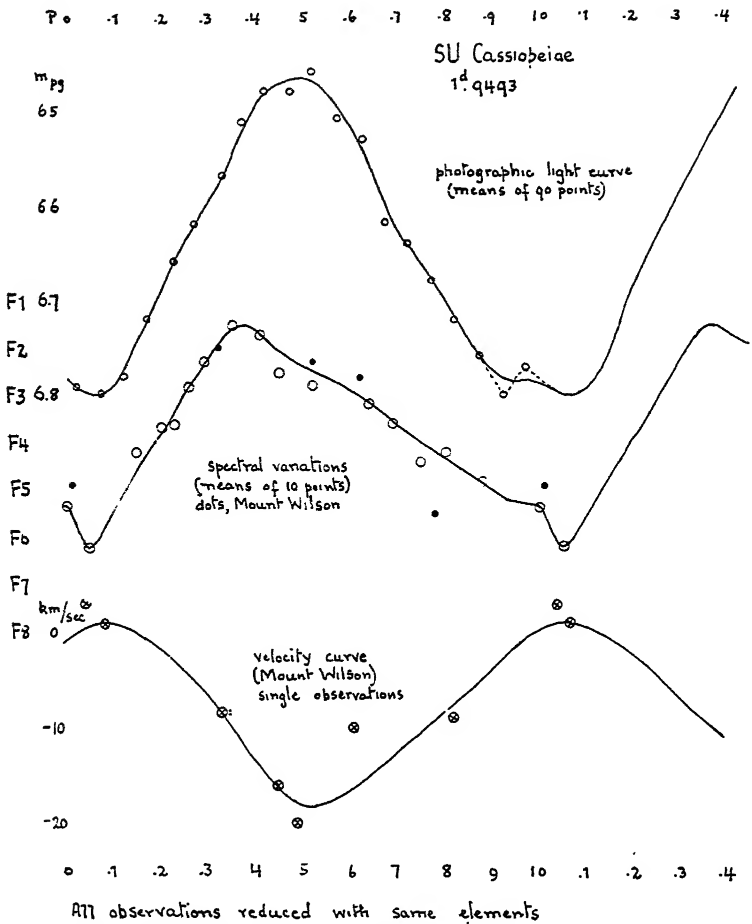


FIG. 6. Mean photographic light curve (above), mean spectral curve (center), and observations of radial velocity (below) for SU Cassiopeiae. Ordinates for the three curves are, respectively, photographic magnitudes, spectral classes on the Henry Draper system, and kilometers a second. Abscissæ are phases, in terms of the period ($1^d.9493$), corrected empirically for a small change of period. In the photographic light curve, the circles represent means of about ninety points each. In the spectral curve, the circles represent means of ten points each; the dots represent means of the Mount Wilson spectral classes, reduced to the Harvard system. Individual observations are given for the velocity curve.

The light curve of AL Virginis is normal for the period of ten days, as it shows almost complete symmetry. The spectrum, derived from Harvard objective prism plates, is about F5; unfortunately the material is not adequate for a determination of the curve of spectral variation.

d. The Intermediate Variable UW Librae. There has been a general opinion that the maximum period of galactic Cepheids is about forty-five days. A few galactic stars of longer period, such as CG Sagittarii (period 64.1 days) and DN Arae (period 82 days), may be regarded as

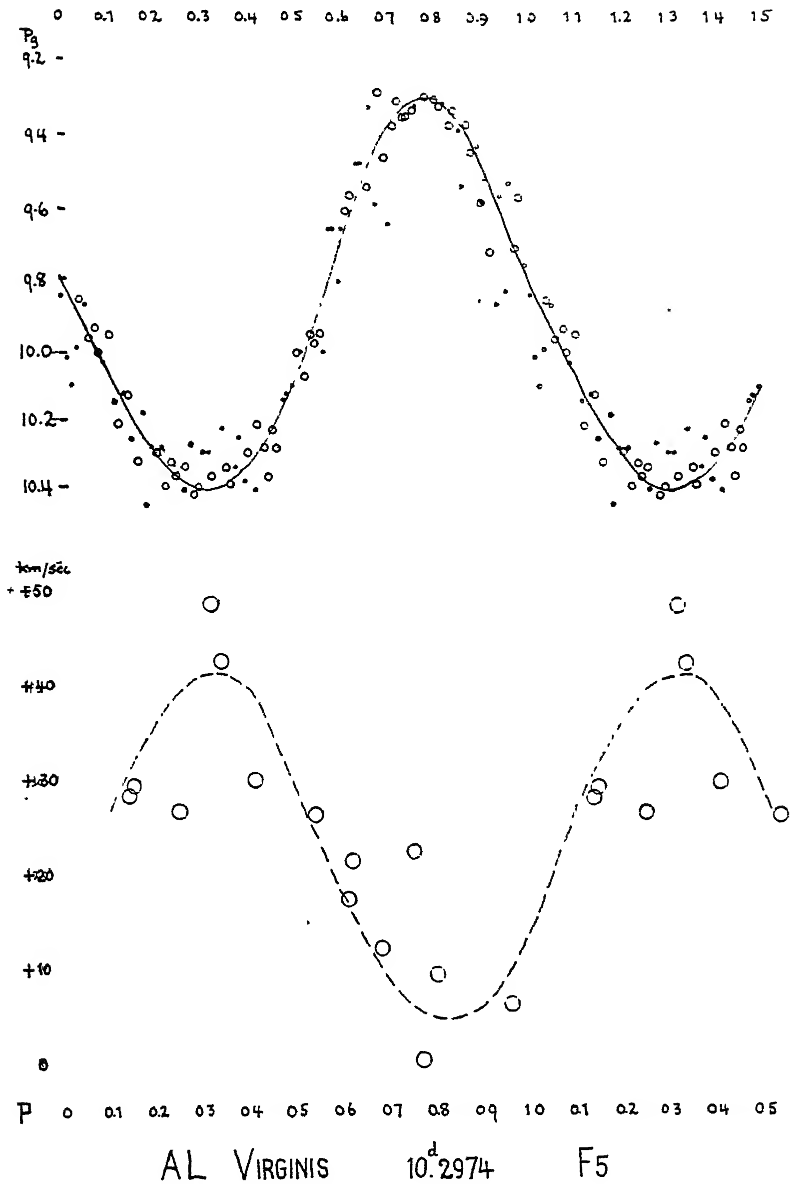


FIG. 7. Mean photographic light curve (above) and velocity curve for the Cepheid variable AL Virginis. Ordinates are photographic magnitudes (above) and kilometers a second; abscissae are phases, in terms of the period ($10^d.2974$). In the mean light curve, circles and dots indicate means of more, and less, than ten observations, respectively. Individual observations (Joy, Mount Wilson) are plotted in the velocity curve (corrected empirically in phase by the data of Fig. 8). Both light and velocity curves are arbitrarily drawn.

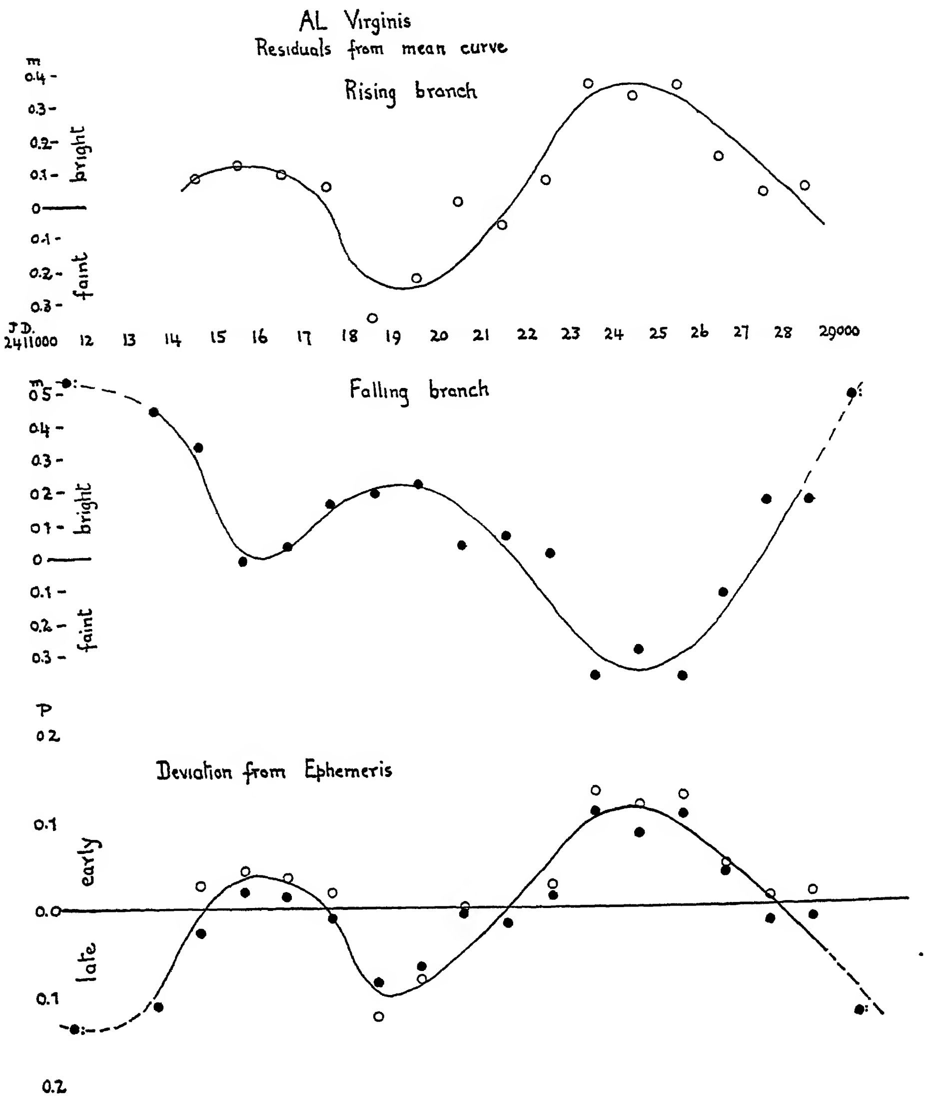


FIG. 8. Deviations of AL Virginis from a uniform ephemeris (see text for description of method). Ordinates are differences of magnitudes (first two curves) and difference of phase (bottom curve) plotted against time in thousands of days. The two upper curves show, respectively, the mean residual on the rising and the falling branches of the light curve; the bottom curve, the deduced deviation from ephemeris (symbols the same as for the two upper curves).

possibly Cepheids also; and in the Magellanic clouds there are several stars, which appear to be Cepheids on the basis of their luminosity, with periods between sixty and a hundred and twenty days.

The star UW Librae has a period of $84^d.73$, as determined from 977 estimates made by Mrs. Helen L. Thomas. It therefore raises an inter-

esting point. Its spectrum is of Class K5. If it is a Cepheid, its absolute photographic magnitude should be -4.32, and this, with its observed median magnitude of $10^m.7$, corresponds, if the star is unobscured, to a distance of 10,000 parsecs. It is then more than six thousand parsecs above the galactic plane, an unduplicated distance for a classical Cepheid, though there are numerous cluster-type variables as far from the plane. The galactic latitude and longitude are $+39°$ and $303°$.

If, on the other hand, the star is a long-period variable, as seems more probable, it has an unusually short period, small range, and early spectrum. It possesses, however, a feature common to many long-period variables—a small but definite variability of the light curve. Observations of the radial velocity of the star are lacking. These, and a determination of proper motion, might assist in assigning it to one or other of the two classes of variables.

e. Three Long-Period Variables. The long-period variables have always been regarded as dangerous and unprofitable subjects for photographic photometry. Their redness makes estimates of their magnitudes more uncertain than those of other stars, both because of the difficulty of finding comparison stars of suitable color, and because their images are of rather unusual quality, especially on plates taken with long-focus cameras.

In the program of photographic study now being initiated, several long-period variables will in general be investigated in one field, on the same series of plates, and by the same measurer. It will thus be possible to note whether a peculiarity of brightness, shown by one star on a particular occasion, is shared by other long-period variables measured on the same plate.

Fig. 10 illustrates the method by comparing the variations of three long-period variables which are all situated in the same field. One is a well-known star; one had been announced as variable but had not been named; the third, though a known variable, was supposed to be irregular in behavior. The observations of all three variables, over the same interval of time, are represented in the figure. All the observations are by Mrs. Helen L. Thomas.

The star RS Librae is of fairly regular behavior. The average deviation of all the observed photographic maxima (not only those represented in the figure) is about one-tenth of a magnitude; that of the maximum magnitudes of SX Librae, on the other hand, is about eight-tenths of a magnitude. The variability of the maximum magnitude of 293.1930 Librae is intermediate.

In the three light curves here illustrated we have an example of the general rule that the light curves that are most variable in maximum brightness are associated with the longest periods. No conclusions can rightly be based on three stars, but when the results of an extensive survey are available it will be possible to relate a numerical measure of the variability of maximal brightness to period and spectrum. It will

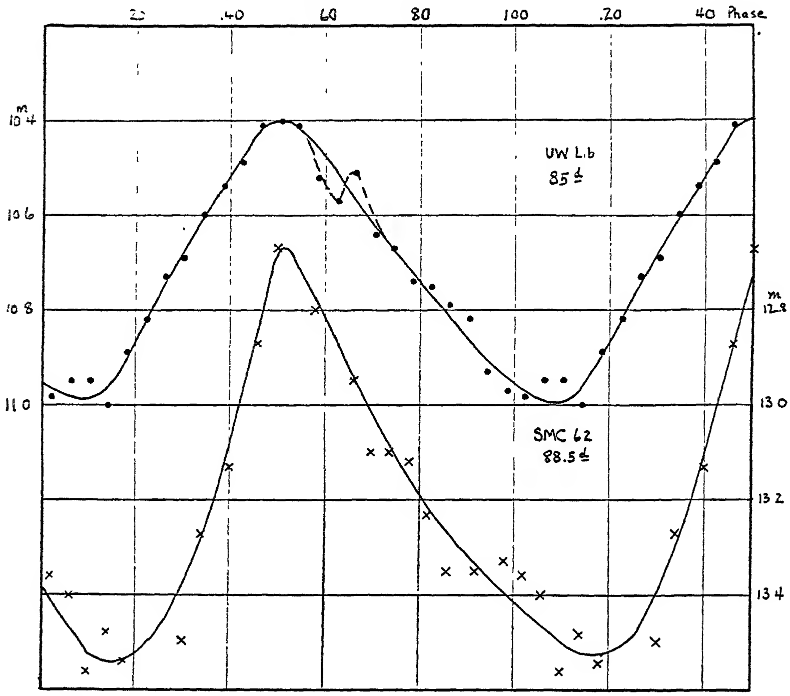


FIG. 9. Mean photographic light curve of UW Librae (above). Below is placed, for comparison, the mean light curve of a Magellanic Cepheid of comparable period (reproduced by kind permission of Miss Hoffleit, who derived it). Ordinates and abscissae are photographic magnitudes and phases in terms of the period (UW Librae, 85 days; Magallenic Cepheid, 88.5 days).

be important to examine whether there are any residual changes that cannot be attributed to small differences in effective temperature at different maxima of the same star. The known association of long period with late spectrum makes this idea very plausible.

SUMMARY

We have presented a program for the systematic study of variable stars. The actual results that have been included are brought out for purposes of illustration. First it is desired to obtain a measure of the precision of the observational methods employed; this is effected by

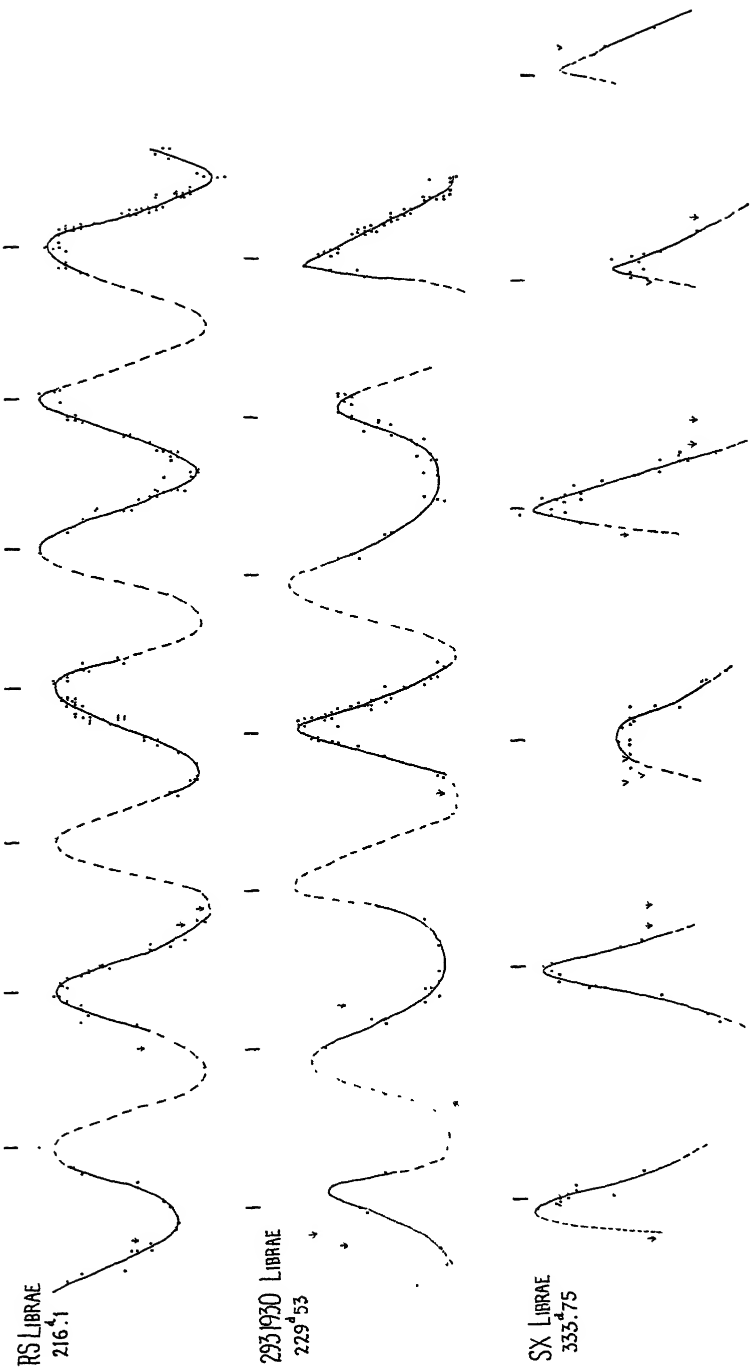


Fig. 10. Variations of three long-period variables over an interval of about two thousand days. The observations are simultaneous; the ordinates are on comparable scales, the stars having ranges between three and six magnitudes. The periods are, respectively: RS Librae, $216^d.1$; 293.1930 Librae, $229^d.53$; SX Librae, $333^d.75$.

using the well-known eclipsing stars as standards. Secondly, the type of result that can be obtained concerning the light curves and their variations, and (in fortunate cases) the spectra and their variations, has been illustrated by a few typical examples. Lastly, a general method for the comparative study of red variables of long period has been suggested.

Reference to the cataclysmic variables has not been made in the foregoing. The most conspicuous of the cataclysmic variables—the supernovæ—are to be discussed in later papers by two of my colleagues. These spectacular irregular variables are, however, an important section of the work of the bureau for the investigation of bright variable stars, and in many ways it is for such stars that a continuous survey will confer the greatest benefits.

The program to which reference has been made is only in its early stages. If, however, it can succeed in furnishing definitive elements and light curves even for a thousand of the brighter variables, there is no doubt that variable star astronomy, both observational and theoretical, will be notably enriched.

STARS WITH EXTENDED ATMOSPHERES

OTTO STRUVE

Director, Yerkes Observatory, University of Chicago

(Read February 17, 1939, in Symposium on Progress in Astrophysics)

I. Departures from Thermodynamic Equilibrium

It has for a long time been considered probable that the emission lines of Be and Ae stars are produced in tenuous shells or rings which surround the stars and which may sometimes extend to very considerable heights above the normal photospheres. In the novæ and the P Cygni type stars we have direct observational evidence of the continuous outflow of very considerable masses of gas which must remain somewhere in space, forming a more or less symmetrical shell around the star. In some Be stars we observe sharp absorption lines of hydrogen which are practically free of Stark effect and apparently are produced in the outermost layers of these shells. These lines are occasionally very strong—showing that the amount of gas in the line of sight is by no means negligible compared with that which we usually find in a normal reversing layer. Hence, we are justified in describing such nebulous envelopes as extended stellar atmospheres. They produce absorption lines in the continuous spectrum of the star, they re-emit the absorbed light and thereby give rise to the emission lines of Be stars and they occasionally produce absorption lines in the continuous spectra of other stars which may happen to be in the same line of sight.

The Balmer absorption lines are most frequently observed, but a fairly large number of stars show in addition the stronger absorption lines of Fe II, and in a few isolated cases—like the one of recent fame, γ Cassiopeiæ [1]—a large number of strong, sharp lines are seen which unmistakably come from the same tenuous atmosphere which produces the Balmer lines. A casual inspection of these spectra reveals nothing unusual. But a more detailed study brings to light a number of extremely interesting anomalies. Perhaps the most significant and puzzling of these is the weakness or even complete absence of the line Mg II 4481. In normal stellar spectra showing Fe II, the magnesium

[1] R. B. Baldwin, *Ap. J.* in press.

line is always very conspicuous. I believe there is no exception to this rule. But in the peculiar spectra of extended atmospheres the lines of Fe II are often strong, while Mg II is invisible.

Before we attempt an explanation, we may profitably investigate whether there are any other peculiarities in stellar spectra which are associated with the Mg II line. Such peculiarities are, indeed, quite numerous. One of the most interesting is the strange behavior of this line in the remarkable binary system ϵ Aurigæ. This double star consists of two components. One is a supergiant of type F5 the spectrum of which is normal for its class and luminosity, and, of course, contains a strong Mg II line. The other is a supergiant of very large size (its radius is about 3000 times larger than the radius of the Sun), the light of which is largely concentrated in the infra-red part of the spectrum, but the absorption spectrum of which is also that which we usually associate with class F5. It has been pointed out elsewhere [2] that the evidence is practically complete that the line spectrum of the infra-red component is excited by the ultra-violet radiation of the yellow F5 component, and that this phenomenon accounts for the discrepancy between effective temperature and line spectrum of the infra-red component. Now, the most striking anomaly in the line spectrum of the infra-red component is the extreme weakness of the line Mg II 4481. This was discovered many years ago, long before there was any hint of the true explanation.

A similar case of unexpected weakness of Mg II occurs in the star 17 Leporis [3] which I observed several years ago. The observations of this star give us another useful clue: Si II 4128 and 4131 are also unusually weak, although both they and Mg II should normally dominate this spectrum, which is full of Fe II, Ti II, Cr II, Sc II, etc.

These cases could easily be amplified but we shall content ourselves with only one more example: the famous B5 component of β Lyræ, long believed to represent the spectrum of the second component of this binary, but now definitely attributed to absorption in the nebulous shell which gives rise to the emission lines,[4] contains only the weakest trace of Mg II, although the He I lines are extraordinarily strong.

It is quite clear that the anomalous behavior of Mg II can not be due to differences in ionization or excitation. It might be due to differences in abundance, but it would certainly be strange if only Mg II and Si II should be affected, while Fe II and many other elements are perfectly normal. Moreover, experience has taught us to be cau-

[2] Kuiper, Struve and Strömgren, *Ap. J.*, **87**, 570, 1937.
[3] Struve, *Ap. J.*, **76**, 86, 1932.
[4] Struve, *The Observatory*, **57**, 265, 1934.

tious where abundance is considered. It is definitely preferable to investigate other possibilities first. What is it that is common to Mg II 4481 and to the two Si II lines, and that is not shared by Fe II, Ti II, Cr II, Sc II, etc.?

The answer is: the lower energy levels of the Mg II and Si II lines are not metastable, while the lower levels of all observable (with the Yerkes spectrograph) lines of Fe II, Ti II, etc. are metastable.

When we consider this result in the light of the fact that in *all* cases of abnormal weakness of Mg II and Si II the lines were already from other evidence believed to originate in a nebula (β Lyræ), in an expanding shell (17 Leporis), or in a tenuous infra-red star excited by the radiation of a distant hot companion, we begin to feel that we have struck upon the true explanation; when the source of the radiation is distant and the so-called dilution factor markedly differs from one, atoms will accumulate in the ground level of the atom and in its various metastable levels. Ordinary levels will be rapidly depopulated, and Rosseland's theory of cycles suggests that this process of depopulation proceeds in exact proportion to the dilution factor.

In order to test this hypothesis we have recently examined the spectra of a number of peculiar B stars in which emission lines give convincing evidence of the existence of extended atmospheres or shells. The spectrum of He I is of particular interest in this connection, because some of its levels are metastable, while others are not. The results of these investigations have been fully described in a paper by Struve and Wurm,[5] and important additional evidence has recently been secured by R. B. Baldwin from a study of the spectrum of γ Cassiopeiæ.[1] In all stars which show sharp absorption lines of H centrally superposed over a broad emission line we find the lines 3888 and 3965 stronger than normal. These lines have metastable lower levels (2^3S and 2^1S). The other levels are not metastable and the lines originating from them are weak. The recognition of this phenomenon has already produced valuable results; it has enabled us to determine, roughly, the dilution of the radiation and hence the distance of the photosphere from the gas in which the lines are formed. The discussion by Struve and Wurm leads to a value for the radius of the atmosphere:

$$R_{\text{shell}} = 5R_*.$$

II. The Theory of Cycles

1. It has usually been assumed, sometimes tacitly, that the populations of the various excited levels in a gas in the reversing layer of a

[5] *Ap. J.*, 88, 84, 1938.

star may be computed by means of the Boltzmann formula:

$$\frac{n_k}{n_1} = \frac{g_k}{g_1} e^{-h\nu_{1k}/kT}, \tag{1}$$

and that the state of ionization is given by Saha's formula:

$$\frac{n_{r+1}}{n_r} P_e = \frac{u_{r+1}}{u_r} 2 \frac{(2\pi m)^{3/2}(kT)^{5/2}}{h^3} e^{-\chi_r/kT}, \tag{2}$$

where g_k and g_1 are the statistical weights of the excited state (k) and of the ground state; h is Planck's constant ($h = 6.55 \times 10^{-27}$ erg sec); k is Boltzmann's constant ($k = 1.37 \times 10^{-16}$ erg/degree); T is the temperature; ν_{1k} is the frequency of the line which corresponds to the transition $1 \rightarrow k$; P_e is the pressure of the free electrons; χ_r is the ionization potential in volts; u_{r+1} and u_r are the partition functions of the atom in the $(r + 1)$ and the (r) stages of ionization; m is the mass of the electron ($m = 9 \times 10^{-28}$ gr).

As we have seen in section I, we have, however, a large amount of evidence that in some stars, at least, the reversing layers are very greatly extended, and the question arises whether conditions of thermodynamic equilibrium may be appropriately used in these cases.

For example, in the star ζ Aurigæ the observations during the eclipse of the small B-type star by the atmosphere of the large K-type star clearly show that this atmosphere extends to at least $0.2R_K$,[6] where R_K designates the radius of the K-type component. A similar atmosphere was found for the infra-red component of ϵ Aurigæ, where it was estimated to be about 3×10^8 km in height, while the radius of the star is 2×10^9 km. This would make the thickness of the atmosphere approximately $0.15R_I$.

The preliminary results by Goedicke for VV Cephei give even greater heights.[7] Finally, it is considered probable that very tenuous atmospheres having a height of about 10^8 km surround some peculiar A and B type stars, such as ζ Tauri, ϕ Persei, etc.

2. It is necessary to examine whether for such extended atmospheres formulæ (1) and (2) are still valid. We, therefore, introduce the idea of "dilution" of radiation. In thermodynamic equilibrium, where every elementary volume of gas receives from all sides the amount of radiation which is in equilibrium with its temperature T, we designate the "dilution factor" as unity. If the energy-density of the radiation is less than that which corresponds to the temperature T we

[6] Christie and Wilson, *Ap. J.*, 81, 426, 1935.
[7] *Pub. Am. Astr. Soc*, 9, 122, 1938.

define the dilution factor as follows:

$$\beta = \frac{\text{Energy density of radiation}}{\text{Energy density at equilibrium with } T}.$$

In practice this means that, although the distribution of energy as a function of frequency may correspond to a temperature T, the density of this radiation is not that given by Planck's formula, but is equal to the latter times the factor β.

At the outer surface of a star, if one is permitted to speak of such a surface, $\beta = 0.5$, because as seen from the elementary volume of gas only one hemisphere is occupied by the luminous photosphere of the star. It is generally assumed, however, that in reality the absorption line is produced in a region where the difference in outgoing and incoming radiation is relatively small, so that one may assume $\beta = 1$.

However, when we are dealing with very tenuous extended atmospheres this is no longer true. These atmospheres emit little or no light in the continuous spectrum. Hence, in their lower regions $\beta = 0.5$. In the higher strata β is even smaller.

It is easy to see that

$$\beta = \frac{\omega}{4\pi} = \frac{1}{2}\left\{1 - \sqrt{1 - \frac{R^2}{r^2}}\right\}. \tag{3}$$

If we expand the bracket in a series we obtain the usual formula:

$$\beta = \frac{\omega}{4\pi} = -\frac{1}{2}\left\{\left(1 - \frac{1}{2}\frac{R^2}{r^2} - \frac{1}{8}\frac{R^4}{r^4}\cdots\right) - 1\right\} \approx \frac{R^2}{4r^2}, \tag{4}$$

which is good when $r \gg R$.

It is of interest to compute β for $r = 1.2R$, which corresponds to the case of ζ Aurigæ. From (3) we find

$$\beta = 0.22.$$

We see that even for a relatively small thickness of atmosphere β can become appreciably smaller than one.

3. We shall now investigate the problem of the distribution of the atoms of an element among its various energy levels. This problem has been discussed by Rosseland,[8] by Ambarzumian,[9] by Woolley,[10] by Struve and Wurm,[11] and by others. In most of the former investigations one or even both of two assumptions were made: (a) β is very

[8] *Ap. J.*, 63, 218, 1926; *Astrophysik auf Atomtheoretischer Grundlage*, p. 224, 1931.
[9] *Poulkovo Obs. Circ. No. 6*, 1933.
[10] *Monthly Notices R. A. S.*, 94, 631, 1934.
[11] *Ap. J.*, 88, 93, 1938.

small and (b) one of the energy levels is metastable. In reviewing the results of Rosseland, Ambarzumian, etc., we shall avoid these restrictions and consider the problem in a more general form. In principle, the problem is completely solved when all transition probabilities and the value of β are known, no matter what the number of levels. But in practice the formulæ become exceedingly long when more than three, or in special cases four, levels are considered. This restriction is not serious when we are concerned with general properties of the solutions. When more levels must be taken into account a numerical computation, similar to that carried out by Struve and Wurm [11] for He I, may always be resorted to.

We start with an atom having an arbitrary number of levels, the populations of which we designate as $n_1, n_2, n_3 \cdots$. Let us designate the numbers of transitions per atom per second from level i to level k as a_{ik} and from level k to level i as a_{ki}. Let us next define, for simplicity of writing:

$$a_{ii} = -(a_{i1} + a_{i2} + \cdots + a_{i(i-1)} + a_{i(i+1)} + \cdots a_{ik} + \cdots). \quad (5)$$

If we then equate the number of atoms per second which leave each state and the number which arrive in that state, *i.e.*

$$n_1 a_{1i} + n_2 a_{2i} + \cdots = n_i(a_{i1} + a_{i2} + \cdots) = -n_i a_{ii}, \quad (6)$$

we obtain a system of linear equations of the form

$$\sum_k a_{ki} n_k = 0. \qquad i = 1, 2, 3, \cdots.$$

Hence the determinant

$$D = \begin{vmatrix} a_{11} & a_{21} & a_{31} & \cdots \\ a_{12} & a_{22} & a_{32} & \cdots \\ a_{13} & a_{23} & a_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix} = 0 \quad (7)$$

and the unknowns $n_1, n_2 \cdots$ are proportional to the minors of this determinant:

$$n_1 = C \begin{vmatrix} a_{22} & a_{32} & \cdots \\ a_{23} & a_{33} & \cdots \\ \cdot & \cdot & \cdot \end{vmatrix} \quad (8)$$

$$n_2 = C \begin{vmatrix} a_{12} & a_{32} & \cdots \\ a_{13} & a_{23} & \cdots \\ \cdot & \cdot & \cdot \end{vmatrix} \text{etc.}$$

The coefficients a_{ik} may be expressed, in the usual way, by means of the Einstein probability factors:

$$\begin{aligned} a_{ik} &= B_{ik}\beta\sigma_{ik}\rho_{ik} \\ a_{ki} &= A_{ki} + B_{ki}\beta\sigma_{ik}\rho_{ik}, \end{aligned} \tag{9}$$

where $\sigma_{ik} = \dfrac{8\pi h\nu_{ik}^3}{c^3}$ and $\rho_{ik} = \dfrac{1}{e^{h\nu_{ik}/kT} - 1}$.

We recall the standard relations between the Einstein factors:

$$\begin{aligned} A_{ki} &= B_{ki}\sigma_{ik}; \qquad & A_{ki} &= B_{ik}\frac{g_i}{g_k}\sigma_{ik} \\ B_{ik} &= B_{ki}\frac{g_k}{g_i} \qquad & B_{ik} &= \frac{g_k}{g_i}\frac{1}{\sigma_{ik}}A_{ki}. \end{aligned} \tag{10}$$

When $\beta = 1$, formulæ (8) automatically give the Boltzmann distribution, in accordance with (1).

In many astrophysical problems it is not necessary to consider the induced emissions. For example, when we have only two states and when $\beta = 1$, we see at once that in place of (1) we obtain

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}\rho_{12}. \tag{11}$$

The error arises in writing $\rho_{12} = 1/e^{h\nu_{12}/kT}$ in place of $1/(e^{h\nu_{12}/kT} - 1)$. Whenever these two expressions are sufficiently close to each other, the neglect of the induced emissions is justified. This condition remains in force when $\beta < 1$.

Although it is not essential that we neglect the induced emissions we shall do so throughout this paper, since otherwise the formulæ become very complicated.

4. The theory of Ambarzumian [9] has shown that for *three* states we have the following expression, when $\beta \ll A_{21}/A_{31}$, and when $A_{21} \ll A_{31}$:

$$\frac{n_2}{n_1} = \beta\frac{A_{31}}{2A_{21}}\rho_{13}. \tag{12}$$

When $\beta \gg A_{21}/A_{31}$, Ambarzumian finds, in accordance with Rosseland:

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}e^{-h\nu_{12}/kT}. \tag{13}$$

Formula (12) shows that for the conditions for which it is valid the ratio n_2/n_1 is inversely proportional to the transition probability A_{21}. The question arises whether in a multiplet the populations of the multiple level will be altered in any way by the downward transitions.

It is not immediately obvious what the answer will be and we shall, therefore, consider first a normal multiplet with Russell-Saunders coupling for which the sum rule of Burger and Dorgelo is valid.

We consider four states, among which states 2 and 3 are supposed to be members of the multiplet. We exclude transitions between these two levels ($a_{23} = a_{32} = 0$). Then it is easily seen that

$$n_1 = C \begin{vmatrix} a_{22} & 0 & a_{42} \\ 0 & a_{33} & a_{43} \\ a_{24} & a_{34} & a_{44} \end{vmatrix} \tag{14}$$

$$n_2 = -C \begin{vmatrix} a_{12} & 0 & a_{42} \\ a_{13} & a_{33} & a_{43} \\ a_{14} & a_{34} & a_{44} \end{vmatrix} \tag{15}$$

$$n_3 = C \begin{vmatrix} a_{12} & a_{22} & a_{42} \\ a_{13} & 0 & a_{43} \\ a_{14} & a_{24} & a_{44} \end{vmatrix} \tag{16}$$

Remembering that the sum rule gives $A_{42} = (g_2/g_3)A_{43}$ and $A_{21} = A_{31}$, we find after substitution of the Einstein coefficients

$$n_3/n_2 = g_3/g_2. \tag{17}$$

The numbers of atoms in the multiplet levels remain proportional to their statistical weights, without any restriction on β or on A.

It is important to realize that this result depends upon the sum rule. When this rule is not obeyed, and when $A_{21} \neq A_{31}$ marked anomalies in the ratios n_3/n_2 are bound to occur.

Consider, for example, an intersystem multiplet. Some transitions may be completely forbidden, while others are present. Thus, the ground level of the Mg I atom 3^1S_0 combines with only one of the three triplet levels, $3^3P_1{}^o$, giving the line λ 4571. The other two transitions are forbidden because for one j would change by 2, while for the other it would jump from 0 to 0. It is clear that in this case the triplet levels $3^3P_0{}^o$ and $3^3P_2{}^o$ will behave as metastable levels, while 3^3P_1 will behave as an ordinary level. We shall have, approximately,

$$\left.\begin{aligned} n_2(3^3P_0{}^o)/n_1 &= \frac{g_2'}{g_1} e^{-h\nu/kT} \\ n_2(3^3P_1{}^o)/n_1 &= \frac{g_2''}{g_1} \beta e^{-h\nu/kT} \\ n_2(3^3P_2{}^o)/n_1 &= \frac{g_2'''}{g_1} e^{-h\nu/kT} \end{aligned}\right\} \tag{18}$$

and, therefore,

$$\frac{n_2(3^3P_1^o)}{n_2(3^3P_0^o)} = \beta. \tag{19}$$

This effect should be easily observable when we consider a normal multiplet arising from the 3^3P^o levels. For example, we should expect that in the green multiplet λλ 5184, 5172, 5167 ($3^3P^o - 4^3S$) the middle line will be weakened by the factor β, relative to the other two lines. The same should be true of the line λ 3832 in the multiplet ($3^3P^o - 3^3D$) which also contains the lines λλ 3828 and 3829.

Among the many spectra with two valency electrons which have essentially the same structure as the spectrum of Mg I, is that of Ca I. Here the situation is exactly analogous. The line λ 6573 combines the 4^1S_0 term and the $4^3P_1^o$ term. The two other ^3P terms are metastable. Hence the effect of β should be pronounced in the line λ 4435 ($4^3P_1^o - 4^3D$) and should be absent in the other two members of this triplet, λλ 4425 and 4455. An observational test of the Mg I and the Ca I triplets is entirely possible.

In He I conditions are less favorable. The lowest level of the triplet system is 2^3S, which is completely metastable.[12] The three 2^3P levels lie slightly higher and combine with the 2^3S level (λλ 10830, 10829). It is possible that one of these levels, $2^3P_1^o$, combines with the ground level, 1^1S_0, giving the line λ 591.6, but the evidence is not complete. We should expect that Russell-Saunders coupling would be much more nearly realized in He I than in the more complicated spectra. Hence there is reasonable doubt that the identification of the line λ 591.6 with He I is correct.[13] An observational test based upon a comparison of the He I multiplet components of λ 5876 seems out of the question, because the transitions to 2^3S will doubtless dominate the populations of the 2^3P levels, and also because the two components arising from the levels $2^3P_2^o$ and $2^3P_1^o$ (one of which certainly does not combine with 1^1S_0 while the other may, because of the selection rule for j) cannot in practice be separated.

There are numerous other cases among the astrophysically important spectra in which the theory predicts interesting results. For example, the case of H deserves careful study. The $2S_{1/2}$ level is metastable, because of the selection rule for the azimuthal quantum number. The $2P_{1/2}$ level, on the other hand, is not metastable, although it exactly coincides with the $2S_{1/2}$ level. Nor is the $2P_{3/2}$ level metastable. Hence

[12] Struve, Wurm and Henyey, *Proc. Nat. Acad. Sciences*, 25, 67, 1939.
[13] See: Bacher and Goudsmit, *Atomic Energy States*, p. 17, 1932.

we should expect that, if β is small, one component of each Balmer line would be relatively weakened, while the other would remain essentially unaltered. The effect in wave length would, of course, be very small. For $H\alpha$ it is only about 0.16A and for the other Balmer lines the displacement is even smaller. There is little hope that we shall soon be able to measure this small difference, which, of course, would have to be established by relating the Balmer wave lengths to those of other stellar absorption lines. There is, however, much more hope that we shall soon be able to test the possible influence of the metastability of the $2S_{1/2}$ level by comparing the Balmer lines with the corresponding Paschen lines, in stars having outer shells—ζ Tauri, ϕ Persei, etc.—and in normal supergiants, β Orionis, α Cygni. It is important that supergiants be selected for this comparison, so that the influence of Stark effect may be ignored.

From the point of view of the theory the H lines present a particularly interesting case. The $2S_{1/2}$ level has a practically infinite lifetime if there are no outside perturbations. Bethe [14] states that the lifetime is several months. But in the presence of electrical fields the $2S$ level gradually becomes merged with the $2P$ levels and the lifetime becomes very much shorter. Henyey [12] has made a computation upon the basis of the theory by Bethe and has concluded that the lifetime becomes of the order of 10^{-8} sec if $n_e = 10^{10}$ cm^{-3}. In general, we may write

$$\tau_{2S} = 4 \times 10^5 n_e^{-4/3}. \tag{20}$$

The electron density required to lose completely the character of a metastable level is, thus, similar to that of the solar chromosphere. One may estimate in the outer shells of stars $n_e = 10^6$ cm^{-3}. Hence the $2S$ level should be essentially metastable, and formulæ (12) or (13) should apply, depending upon the value of β.

The He II spectrum is not suitable for an observational test, in spite of the large fine structure of λ 4686, because the lower level is not metastable. The lines originating from the second energy level all lie in the extreme ultra-violet region of the spectrum.

5. We shall now continue our discussion of the theory of cycles, without making the usual assumptions that β is small and that one of the levels is metastable. The reason for this extension of the theory is that in many spectra the values of A_{21} are not all the same (as they are in a normal multiplet). For example, in the case of Ca I, we might compare the multiplets $(4^1P^o - 4^1D)$ λ7326; $(3^1D - 4^1F^o)$ λ4878 and $(4^3P^o - 4^3D)$, λλ4425, 4435, 4455, etc. The lower level of the first

[14] *Handbuch der Physik*, **24**, part 1, p. 452.

combines with the ground level by means of the resonance line λ4227. The lower level of the second multiplet is completely metastable, while that of the third multiplet is partly metastable and partly connects with the ground level by means of the relatively weak inter-system line λ6573. It is clear that the first and third multiplets will behave differently when $\beta \neq 1$, because A_{21} is not the same, although in neither case can we speak of metastability.

We shall use formulæ (14) and (15) and neglect, for simplicity, possible differences in the statistical weights. After substituting the Einstein coefficients we obtain

$$\frac{n_2}{n_1} = \frac{-A_{21}\beta\rho_{12}(A_{41} + 3A_{42}) - A_{41}\beta\rho_{14}A_{42}}{-(A_{21} + A_{42}\beta\rho_{24})(A_{41} + 3A_{42}) + 3A_{42}{}^2\beta\rho_{24}}. \tag{21}$$

It is now convenient to introduce

$$\left.\begin{aligned} A_{41} &= \gamma A_{42} \\ A_{41} &= \alpha A_{21} \end{aligned}\right\}. \tag{22}$$

We then have

$$\frac{n_2}{n_1} = \frac{-\beta\rho_{12}\left(1 + \frac{3}{\gamma}\right) - \frac{\alpha}{\gamma}\beta\rho_{14}}{-\left(1 + \frac{\alpha}{\gamma}\beta\rho_{24}\right)\left(1 + \frac{3}{\gamma}\right) + \frac{3\alpha}{\gamma^2}\beta\rho_{24}}$$
$$= \frac{\beta\rho_{12}\left(1 + \frac{3}{\gamma}\right) + \frac{\alpha}{\gamma}\beta\rho_{14}}{\left(1 + \frac{3}{\gamma}\right) + \frac{\alpha}{\gamma}\beta\rho_{24}}. \tag{23}$$

The only simplifications which we have made thus far are (*a*) the neglect of induced emissions, (*b*) the neglect of statistical weights. We now introduce the further simplification

$$\rho = e^{-h\nu/kT}. \tag{24}$$

Then

$$\rho_{14} = \rho_{12}\rho_{24}, \tag{25}$$

and we finally have

$$\frac{n_2}{n_1} = \beta\rho_{12}\left[\frac{(\gamma + 3) + \alpha\rho_{24}}{(\gamma + 3) + \alpha\beta\rho_{24}}\right]. \tag{26}$$

When $\beta = 1$ we have $n_2/n_1 = \rho_{12}$, which is identical with Boltzmann's formula (1), except for the statistical weights which we have neglected. When β is small, we have two limiting cases: (*a*) when $\alpha\beta\rho_{24} \gg (\gamma + 3)$

we have

$$\frac{n_2}{n_1} = \rho_{12}, \tag{27}$$

which is Rosseland's formula (13), without the statistical weights. (*b*) when $(\gamma + 3) \gg \alpha\beta\rho_{24}$,

$$\frac{n_2}{n_1} = \beta\rho_{12} + \frac{\alpha}{(\gamma + 3)}\beta\rho_{14}, \tag{28}$$

which agrees, essentially, with one of Ambarzumian's formulæ. When the level 2 is not metastable, we may have, approximately, $\alpha \approx \gamma \approx 1$. In that case

$$\frac{n_2}{n_1} = \beta\rho_{12} + \tfrac{1}{4}\beta\rho_{14}. \tag{29}$$

But if the second level is metastable, or if, generally,

$$\frac{\alpha}{(\gamma + 3)}\beta\rho_{14} \gg \beta\rho_{12},$$

we have approximately

$$\frac{n_2}{n_1} = \frac{\alpha}{(\gamma + 3)}\beta\rho_{14}, \tag{30}$$

which is our alternative for Ambarzumian's formula (12).

The general expression (26) is often useful to compute n_2/n_1, when there are no restrictions on β, α and γ. As an example we shall take an arbitrary case which, roughly, corresponds to Ca I at $T = 10{,}080°$. We take the three energy levels

1–2	$\sim$ 2 volts	$\rho_{12} = 0.1$
1–4	$\sim$ 6 volts	$\rho_{14} = 0.001$
2–4	$\sim$ 4 volts	$\rho_{24} = 0.01.$

The following table gives the relative populations for different values of

TABLE 1

VALUES OF n_2/n_1

β	$\alpha = 9$, $\gamma = 3$	$\alpha = 90$, $\gamma = 3$	$\alpha = 900$, $\gamma = 3$
1.	0.1000	0.1000	0.1000
0.1	0.0101	0.0113	0.0217
0.01	0.0010	0.0011	0.0025
0.001	0.0001	0.0001	0.0002

β and of α. The Boltzmann formula gives $n_2/n_1 = 0.1$, in exact agreement with the tabular values for $\beta = 1$. The table also shows

that when α is small the values of n_2/n_1 are almost exactly proportional to β, in agreement with (29). But when α is large, departures appear, so that for relatively large β we have constant n_2/n_1, in accordance with (27) and for decreasing β, a gradual approach to proportionality, according to (30).

6. Another, rather interesting, case remains to be discussed, because it has direct bearing upon the interpretation of astronomical spectra. We ask, what will be the effect upon n_2/n_1 of different values of ρ_{24}? Suppose, for example, that we observe in a Nova two lines, not necessarily of the same element, both of which originate from metastable levels, so that we may, reasonably, assume for both lines:

$$\alpha = 10^6, \qquad \gamma = 1,$$

but that for one line the corresponding ρ_{24} is large, while for the other it is small. How will the two lines behave if β decreases with time as the result of expansion? Evidently

$$\frac{n_2}{n_1} = \beta\rho_{12}\left[\frac{4 + 10^6\rho_{24}}{4 + 10^6\beta\rho_{24}}\right]. \tag{31}$$

The result is shown in Table 2, which gives the values of the expression in brackets as function of β and ρ_{24}. The important thing is, of course,

TABLE 2

VALUES OF $\dfrac{n_2}{n_1\beta\rho_{12}}$

β \ ρ_{24}	10	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.0	1.0
10^{-1}	10.0	10.00	9.96	9.65	7.50	2.80	1.22	1.02	1.00	1.0	1.0
10^{-2}	10.00×10^1	9.96×10^1	9.62×10^1	71.7	20.8	3.41	1.25	1.02	1.00	1.0	1.0
10^{-3}	9.96×10^2	9.62×10^2	7.15×10^2	201.0	25.4	3.49	1.25	1.02	1.00	1.0	1.0
10^{-4}	9.62×10^3	7.15×10^3	2.00×10^3	245.0	25.9	3.50	1.25	1.02	1.00	1.0	1.0
10^{-5}	7.15×10^4	2.00×10^4	2.45×10^3	250.0	26.0	3.50	1.25	1.02	1.00	1.0	1.0
10^{-6}	2.00×10^5	2.44×10^4	2.49×10^3	251.0	26.0	3.50	1.25	1.02	1.00	1.0	1.0
10^{-7}	2.44×10^5	2.49×10^4	2.50×10^3	251.0	26.0	3.50	1.25	1.02	1.00	1.0	1.0
10^{-8}	2.49×10^5	2.50×10^4	2.50×10^3	251.0	26.0	3.50	1.25	1.02	1.00	1.0	1.0
10^{-9}	2.50×10^5	2.50×10^4	2.50×10^3	251.0	26.0	3.50	1.25	1.02	1.00	1.0	1.0
10^{-10}	2.50×10^5	2.50×10^4	2.50×10^3	251.0	26.0	3.50	1.25	1.02	1.00	1.0	1.0

that when the fourth level is very high above the second, n_2/n_1 is always equal to $\beta\rho_{12}$, just as though the second level were not metastable at all; transitions by way of the fourth level become so rare that, in spite of the large value of α, the interchange is solely between one and two. For any given value of β we find that n_2/n_1 is always largest when ρ_{24}

is largest, *i.e.* when the second and the fourth levels are close to each other. But if we examine the trend of n_2/n_1 within each column, we find that the maximum is reached much earlier when ρ_{24} is small than when it is large. It is conceivable that this effect could be observed in the spectra of novæ, where either the forbidden emission lines, or the permitted absorption lines which originate from the metastable levels, may be used to determine n_2/n_1. It is necessary to remember in this connection that we have, arbitrarily, assumed that $\gamma = 1$. In general, this will, of course, not be the case and there should be a correlation between γ and ρ_{24}, but the variety of the spectra is so great, and our present knowledge of transition probabilities so fragmentary, that it is useless to pursue this question further.[15]

We have completely ignored, so far, the effects of collisions. While in nebulæ and in many extended atmospheres collisions play only a relatively insignificant role, we know that they can be important when the energy differences between two levels are small. This problem has been investigated by Wurm,[16] who concludes that in the presence of diluted stellar radiation electron impact excitation will tend to produce larger populations in metastable levels than in normal levels, and that this process will become more and more conspicuous as the electron density is decreased. We shall not discuss in this paper the collisional processes, but merely point out that there is no *intrinsic* difference between excitation by radiation and excitation by collision. Both processes will give rise to similar phenomena, provided the electron density is sufficiently low. The tendency of collisions to bring about a Boltzmann distribution will slow up the process of redistributing the population so that departures which are conspicuous for a certain value of β may not become conspicuous until β has reached a somewhat smaller value, but the principle of the problem is not altered. Interesting discussions of the effect of collisions have recently been given by Menzel[17] and by Bowen and Minkowski.[18]

[15] In a very important paper on Nova Herculis, *Zs. f. Ap.*, **13**, 222, 1937, Grotrian has attempted to explain the early appearance of the forbidden [O I] lines in emission and the late appearance of the forbidden [Fe II] lines, also in emission, by a mechanism which is essentially similar to that which we have here discussed. Our Table 2 shows that Grotrian's explanation will not apply solely because for [O I] ρ_{24} is small, while for [Fe II] it is large. This might produce an earlier maximum in [O I], but this effect is caused by the suppression of [O I] when β is very small and not by a real strengthening of [O I] in the early stages. But in principle it is well possible that α, γ and ρ_{24} combine to produce the observed phenomena.

[16] *Zs. f. Ap.*, **14**, 321, 1937.

[17] *Nature*, **142**, 644, 1939.

[18] *Ibid.*, **142**, 1079, 1939.

III. Nova Herculis

1. As an illustration of the relations discussed in the preceding section we shall consider the spectrum of Nova Herculis 1934. This spectrum has been fully discussed by McLaughlin.[19] The Nova displayed several sets of absorption lines, some of which lasted only a few days. But the absorption spectrum, designated as II by McLaughlin, lasted from the end of December until the latter part of March. This spectrum resembled ϵ Aurigæ; the lines were strong and sharp and there were no violent changes in the Nova during its lifetime. This part of the spectrographic history of the Nova has also been investigated by Stratton,[20] by Grotrian and Rambauske[21] and by others. Hence we possess an unusually complete record of this particular spectrum.

Although we are not certain of the nature of the processes giving rise to a nova absorption spectrum, an investigation of departures from thermodynamic equilibrium is most appropriate. If such departures exist we shall not necessarily be able to decide whether the absorbing shell travels outward as a thin layer or whether the atoms are continuously being replenished from the photosphere.[22] It is to be expected that both processes will be connected with departures from equilibrium conditions. It is, however, clear that in one case the departures should follow from the expansion of the shell which may be computed from the radial velocity, while in the other it must be partly caused by the gradual disintegration of the continuous emission of the expanded photosphere, which culminated in the rapid decrease of the star's brightness in the beginning of April.

We shall discuss only the spectrographic evidence, without making an attempt to use the information for the interpretation of the physical processes involved.

Our first evidence comes from the line Mg II 4481. While the spectrum, as a whole, changed very little in appearance during the latter part of December and all of January, the Mg II line, which was at first exceedingly strong, rapidly faded, so that by January 5 it was quite inconspicuous and by January 11 it had completely disappeared. This was noticed by Guthnick[23] and has been commented upon by McLaughlin.[19] The similarity to the behavior of this line in Be stars (see Section I) is obvious. But we must first investigate whether the

[19] *Pub. Obs. U. of Michigan*, **6**, No. 12, 1937.

[20] *Ann. Solar Phys. Obs. Cambridge*, **4**, 133, 1936.

[21] *Zs. f. Ap.*, **10**, 209, 1935.

[22] See: F. L. Whipple and C. Payne-Gaposchkin, *Harvard Obs. Circ. No. 413*, 1936; *Proc. Nat. Acad. Science*, **22**, 195, 1936.

[23] *Naturwissenschaften*, **23**, 251, 1935.

weakening may perhaps be attributed to a decrease of temperature or to a change in ionization. Both alternatives are excluded by the behavior of other elements. Evidently, no decrease in temperature without appreciable change in ionization can reduce the intensity of the Mg II line by a factor of perhaps 100, without causing a marked increase in the intensities of low-level lines. For example, the Fe I lines of the multiplets $a^5D - z^7P^o$ (λ4216 etc.) and ($a^5D - z^7F^o$) (λ4376 etc.), which are greatly enhanced by α Orionis, are not strengthened in Nova Herculis. A change in ionization is also excluded, because we observe no marked change in the ratio of Fe II/Fe I. There are many other similar considerations, and we are therefore led to believe that the fading of Mg II must be of similar origin as the weakness of this line in 17 Leporis, etc.

We have seen in Section I that Si II 4128 and 4131 are also sensitive to β. McLaughlin lists the lines λλ6347 and 6371, but does not indicate how fast they were fading. He does not list λλ4128 and 4132. These lines are fairly strong in ϵ Aurigæ and α Persei.[24] Grotrian and Rambauske suspect on January 3 a trace of Si II 4131 in a blend and Stratton gives traces of both lines, somewhat blended, on December 27. I believe λ4131 is definitely present on December 28, but it fades rapidly, and is hardly visible on January 1. The other line is hopelessly blended with Fe I. The evidence is not as convincing as in the case of Mg II, but as far as it goes it supports the idea that the radiation becomes rapidly diluted after the formation of shell II.

2. In order to obtain additional data I have examined the spectra of the elements listed by McLaughlin as being represented in Absorption II. The most promising element is Fe I, because it contains fairly strong lines originating from normal and from metastable levels. This is a relatively rare occurrence. Usually the lines observable in the photographic region either all come from metastable levels or all from ordinary levels.

The table of terms of Fe I by Burns and Walters [25] provides the necessary information. The ground level is even, a^5D. The lowest odd level is $z^7D_5^o$, 19,351 cm^{-1}. Hence the low even levels a^5F, 6928 cm^{-1}; a^3F, 11,976 cm^{-1}; a^5P, 17,550 cm^{-1} and a^3P_2, 18,378 cm^{-1} are metastable. These levels give rise to several strong multiplets, for example $a^3F - y^3F^o$, the strongest line of which is λ4045; $a^3F - z^5G^o$, the strongest line of which is λ4384; $a^5P - z^5S^o$, the strongest line of which is λ4282. A casual inspection of the spectra of the Nova shows that

[24] Dunham: Unpublished table of wave lengths.
[25] *Pub. Allegheny Obs.*, 6, No. 11, 1929.

these lines remain uniformly strong over a period of many weeks after the outburst. There are also a number of fairly strong multiplets the lower levels of which are not metastable: $z^7D^o - e^7D$, the strongest line of which is λ4260; $z^7F^o - e^7D$, the strongest line of which is λ4958, etc.

We shall discuss the various multiplets of Fe I in detail.

Multiplet $z^7D^o - e^7D$ (Fe I).

The lower level is normal; it combines directly with the ground level a^5D through the multiplet λλ5166, etc. Since this is an inter-system combination the sum rules cannot be applied and the populations in the levels z^7D^o may not be proportional to their statistical weights. For example, the lower level of λλ4260, 4198, viz. $z^7D_5{}^o$ can combine only with a^5D_4, and this transition is relatively weak, while the lower level of λλ4299, 4235, 4188, viz. $z^7D_4{}^o$, should combine with a^5D_4 by means of a strong transition and with a^5D_3 by means of a weaker transition. In principle the multiplet should be ideal for a study of departures from equilibrium. We shall here examine only the general behavior of the multiplet as a whole.

λ	I Q. No.	Lab. Int.	E. P. v.	Remarks *
4260.488	5–5	35	2.4	Unblended
4271.166	3–4	20	2.4	Blended Fe I 4271.70
4299.252	4–5	18	2.4	Blended Ti I 4300.05
4250.132	2–3	25	2.5	Blended Fe I 4250.79
4235.951	4–4	25	2.4	Unblended
4233.613	1–2	18	2.5	Blended Fe II 4233
4222.223	3–3	12	2.4	Unblended
4210.404 } 4210.337 }	1–1	15	2.5	Unblended
4198.338	5–4	20	2.4	Unblended
4191.439	2–1	15	2.5	Unblended
4187.863 } 4187.787 }	4–3	20	2.4	Unblended
4187.050	3–2	20	2.4	Unblended

* Only prominent blends are listed. If a line is given as unblended we believe that it contributes most to the observed line. The wave lengths are those in the Sun, as listed by Miss C. E. Moore. The blends in the first column are those of unresolved *solar* lines.

The decrease in intensity of λλ4260, 4222, etc. is very conspicuous. The behavior is exactly like that of Mg II 4481. The lines are quite

strong on December 28, but they are already weak on January 1 and disappear by January 5. This weakening seems to be accompanied by a slight differential shift with respect to the neighboring lines. For example, λ4260 seems to shift toward the red with respect to the neighboring lines Fe II 4258 and Cr II 4262, both of which originate from metastable levels. Such a shift is entirely possible, since the different sensitivity of the lines to β will affect the level in the shell at which they are produced. Such differential shifts are conspicuous in P Cygni.

Multiplet $a^3F - z^3G^o$ (Fe I).

λ	I Q No.	Lab. Int.	E P. v	Remarks
4271.776 . .	4–5	35	1.5	Unblended
4250.799.	3–3	25	1.6	Blended Fe I 4250.13
4202.042.	4–4	30	1.5	Unblended
4147.677	4–3	10	1.5	Unblended
4307.914.	3–4	35	1.6	Blended Ti II 4307.89
4325.777.	2–3	35	1.6	Blended Sc II 4325.00

The lower level is metastable. The persistence of these lines way into February is amazing. Even λ4251, which is of similar initial intensity as some of the lines of the preceding multiplet, remains strong when the lines of the latter are completely gone. Only the faintest line, λ4148, which is always weak, does not show the effect clearly. This line is probably weakly present on December 28. It is not observed on January 1 and 2, but a trace is again suspected on January 3 and 4, and perhaps 5. The behavior of λ4251 shows that we are not concerned with a change in turbulence and a consequent alteration of the curve of growth. There remains the possibility that we are dealing with a gradual decrease of temperature and decrease of excitation. This has already been disproved by the fact that the multiplets which originate from the ground level are not strengthened at the same time. They are very weak throughout the life of Absorption II, showing that there is no appreciable decrease in excitation. These multiplets are, of course, unaffected by β. They should behave as do the multiplets with metastable levels, and the fact that they do not become relatively stronger is a convincing point. There are other reasons which disprove a change in excitation.

Multiplet $a^3F - y^3F^o$ (Fe I).

λ	I Q. No.	Lab. Int.	E. P. v.	Remarks
4045.827........	4–4	60	1.5	Unblended
4063.607.......	3–3	45	1.5	Unblended
4071.751.........	2–2	40	1.6	Unblended
4132.069.........	2–3	25	1.6	Unblended
4143.880	3–4	30	1.6	Mostly unblended
4005.256..........	3–2	25	1.6	Probably blended

The lower level is metastable. These lines are very strong and λ4045 persists until the middle of March. The lines are too strong to permit a good comparison with the weaker multiplets having ordinary lower levels.

Multiplet $a^3F - z^5G^o$ (Fe I).

λ	I.Q. No.	Lab. Int.	E. P. v.	Remarks
4383.559..........	4–5	45	1.5	Unblended
4404.763..........	3–4	30	1.6	Unblended
4415.137........	2–3	20	1.6	Blended Sc II 4415.55
4367.914.......	2–2	2	1.6	Blended Ti II
4337.057.......	3–3	10	1.6	Blended Ti II
4294.147 } 4294.050 }	4–4	15	1.6	Blended Ti II 4294.10
4291.475......	3–2	4	1.6	Blended several lines

The lower level is metastable. All but the two strongest lines are affected by blends. The two unblended lines persist at least until the middle of February.

Multiplet $a^5P - z^5S^o$ (Fe I).

λ	I.Q. No.	Lab. Int.	E. P. v.	Remarks
4282.413........	3–2	12	2.2	Unblended
4315.100........	2–2	10	2.2	Blended
4352.745.........	1–2	9	2.2	Blended

The lower level is metastable. The line λ4282 is a conspicuous case of the persisting kind. It remains practically of uniform intensity until after January 5 and may be traced as a faint line considerably later. This case is especially important for three reasons:

(*a*) The lower level has a term value of 17,550 cm^{-1}, which is only 1800 cm^{-1} lower than the lowest odd level listed by Burns and Walters. It is metastable if the z^7D^o level is really the lowest odd level. Professor Russell has been kind enough to inform me that there is no doubt about this matter. It is the lowest of several great triads of odd levels which can be conclusively assigned to electron configurations on Hund's theory. Hence the observations and the spectroscopic theory agree completely.

(*b*) In spite of the small difference in energy between the two levels, collisions do not destroy the observational evidence concerning the metastability of a^5P.

(*c*) The excitation potential of 2.2 volts differs so little from that of multiplet $z^4D^o - e^7D$, namely 2.4 volts, that we can definitely rule out changes in excitation.

Multiplet $b^3P - 19^o$ (Fe I).

Line	I Q. No.	Lab. Int.	E. P. v.	Remarks
4181.766.	2–3	15	2.8	Unblended

Although the level b^3P_2 is not strictly metastable, its term value of 22,838 cm^{-1} is only very slightly higher than those of the three odd levels below it. Moreover, these three levels are septets. Hence it would appear that the transitions between the low even triplets and the odd septets must be very weak, so that, in Professor Russell's words, we should expect "the higher even levels to be very nearly, although not rigorously, metastable, so long as it is a triplet." Observationally λ4182 behaves as though the lower level could be metastable, but the evidence is not conclusive because it is superposed over the strong emission band of the neighboring Fe II lines.

Multiplet $z^7F^o - e^7D$ (Fe I).

The lower level is normal and combines with the ground level through the multiplet $a^5D - z^7F^o$, the strongest line of which is λ4376. This is an extremely interesting case, because the lower level of λ4957.615, $z^7F_6{}^o$ does not directly combine with the ground level, because the largest j number of the latter is 4. Nevertheless, this level is not strictly metastable, because transitions from it to lower even (metastable) levels are permitted. We should, however, anticipate that λ4957.615 will remain stronger than the other lines. Observationally the matter is complicated by blending with λ4957.309 ($z^7F_4{}^o - e^7D_4$), but since the latter is relatively weak a test should be worth while.

Lines	I.Q. No.	Lab. Int.	E P. v.	Remarks
4891.504.........	4–3	50	2.8	Unblended
4920.516.........	5–4	60	2.8	Unblended
4957.615	6–5	60	2.8	Unblended

These lines fade out rapidly, but there is a suspicion that λ4957 disappears a day or two later than λ4920, although they are quite similar in intensity on December 27.9 and 28.0. Observationally, the absence of a strong downward transition from level $z^7F_6{}^o$ produces only a slight effect; the line does not behave as do those which originate from metastable levels. It is not now possible to investigate whether transitions to other even levels, or collisions, are responsible for the fading of λ4957. It will be of interest to study this line in other objects.

Multiplet $b^3P - 16^o$ (Fe I).

Lines	I.Q. No.	Lab. Int.	E. P. v.	Remarks
4175.645...	1–2	10	2.8	Unblended
4156.808..........	2–2	12	2.8	Blended Zr II

The line 4176 is rather faint and the evidence is inconclusive. It is possible that the line persists until about January 5. The evidence is therefore in harmony with that derived from multiplet $b^3P - 19^o$.

Multiplet $b^3G - 24^o$ (Fe I).

Line	I.Q. No.	Lab. Int.	E. P. v.	Remarks
4219.360	4–4	12	3.0	Unblended

The lower level is not strictly metastable, but the same considerations apply to it as to level b^3P. Though not rigorously metastable we expect the transition to the lower septet odd levels to be exceedingly weak. The line persists until about January 5, and behaves therefore intermediate between those having metastable and those having normal levels.

Multiplet $b^3P - x^3D^o$ (Fe I).

Line	I.Q. No	Lab. Int.	E. P. v.	Remarks
4466.564.........	2–3	12	2.8	Blended with $b^3G - 18^\circ$ λ4466.6 and $a^5D - z^7F^o$ λ4466.6

This line persists until January 5 or even 11 and proves beyond doubt that level b^3P does not behave as a normal level. Perhaps the fading is slightly faster than in a line originating from a metastable level.

Multiplet $z^5D^o - e^5F$ (Fe I).

The lower level is normal and connects with the ground level by means of the very strong multiplet $a^5D - z^5D^o$, the strongest line of which is λ3860.

Line	I.Q. No	Lab Int.	E. P. v.	Remarks
4736.783..........	4–5	12	3.2	Unblended

This line disappears rapidly. It is fairly strong on December 28 but is not definitely visible on January 1 or later. We should expect that it would vanish more rapidly than does λ4260, because of the large difference in A_{21}, but the observations do not permit us to answer this question. The spectra are of slightly inferior quality near λ4737.

Multiplet $a^3G - z^3H^o$ (Fe I).

Line	I.Q. No.	Lab. Int.	E P. v.	Remarks
3997.403....... ..	4–5	15	2.7	Blended VII 3997.12
4021.872..........	3–4	12	2.7	Probably unblended
3956.688..........	5–6	12	2.7	Perhaps blended

The lower level is not strictly metastable, but it should behave like b^3P and b^3G. The line 4022 is surrounded by a hazy absorption which may be caused by the confluence of several faint lines. It is difficult to trace the behavior of the Fe I line. Probably it persists until after January 11. λ3957 persists without appreciable change until about January 4. The level thus behaves as do other low even triplets; it is intermediate between normal and metastable levels.

Multiplet $a^5P - x^5P^o$ (Fe I).

Line	I.Q. No	Lab. Int.	E. P. v.	Remarks
4009.719	1–2	10	2.2	Unblended
3977.752...	2–2	12	2.2	Blended VII

The line 4009.72 disappears rapidly, and behaves as though its lower level were permitted. This is the only discrepancy we have found, since a^5P is in reality metastable. That this is true was shown observationally by the multiplet $a^5P - z^5S^o$. I have no explanation to offer. The line does not appear to be badly blended in Dunham's table of α Persei. McLaughlin does not suspect a blend, although he measured it on December 31.

Multiplet $b^3P - w^3D^o$ (Fe I).

Line	I.Q. No.	Lab. Int.	E. P. v.	Remarks
4134.687....	2–3	12	2.8	Unblended
4109.801....	1–1	9	2.8	Blended Nd II?

Multiplet $b^3P - 10^o$ (Fe I).

Line	I Q No.	Lab. Int.	E. P. v.	Remarks
4107.496.........	2–1	12	2.8	Unblended

λ4135 persists until January 5 and supports our former results for b^3P. This line falls on the strong emission of H_δ and no conclusion is possible.

The discussion of the Fe I lines includes all suitable classified lines of Miss Moore's list which have a laboratory intensity of 10 or more. The evidence is, with a single exception, in excellent accord with that derived from the Mg II and Si II lines. It appears that departures from equilibrium conditions are slight on December 28, but that by January 4 or 5 normal levels have become so much depopulated that their lines are no longer observed. No measure of this effect has been made, but we may guess that β decreases by a factor of perhaps 10, and probably not more than 100, within the first week after the appearance of Absorption II.

3. We continue with a brief discussion of other elements.

Fe II.

All classified lines within the range covered by our plates start from metastable levels and all persist until March. The appearance in emission of the forbidden [Fe II] lines shows that by the end of March the forbidden transitions begin to depopulate the metastable levels, and the absorption lines become rapidly weaker. Evidently by the end of March β has become so small that in effect formula (30) will apply, in which the population n_2 depends upon the ratio $\alpha/(\gamma + 3)$. It is

clear that in order to reduce n_2 appreciably β must have become of the order of $1/\alpha$. We have no knowledge of α but we may estimate it to be of the order of 10^6. Hence $\beta \approx 10^{-6}$. This occurred about 12 weeks after the appearance of Absorption II. In the first week we estimated that β had been reduced by a factor of not less than 10^{-2}. Hence, it seems that the process was not linear. Probably the very rapid disintegration of the continuous spectrum in the end of March and beginning of April accelerated the process.

Ti II.

The lines of Ti II persisted longer than those of Fe II and gave the spectrum a very unusual appearance in February and March. The observable lines all come from metastable levels. Hence the persistence is understandable. But it is not obvious why Ti II should persist longer than does Fe II. The effect of ρ_{24} in formula (26) is probably not responsible. Perhaps Ti II is more abundant than in normal stars, but it is tempting to attribute the difference in behavior to $\alpha/(\gamma + 3)$. We are not yet able to answer this question.

Cr II.

All observable lines come from metastable levels and all behave accordingly. One unclassified line, λ4145.81, also behaves as though it originated from a metastable level.

Cr I.

The strong resonance line λ4254.388 is present and persists as long as may be expected.

La II.

This element is interesting because it contains both kinds of lines. $\lambda 3949.062(a^3D_3 - x^3F_4{}^o)$, E. P. 0.4 v. persists, in agreement with its origin from a metastable level. $\lambda 4269.480(z^3F_3{}^o - e^3F_4)$, E. P. 1.8 v. disappears rapidly. Its lower level is normal. This line is supposedly blended with Cr II $4269.292(a^4F_{1\frac{1}{2}} - z^4D_{1\frac{1}{2}}{}^o)$, but the Cr II line cannot be very strong because the two other Cr II lines $4261.935(a^4F_{3\frac{1}{2}} - z^4D_{2\frac{1}{2}}{}^o)$ and $4252.631(a^4F_{2\frac{1}{2}} - z^4D_{2\frac{1}{2}}{}^o)$ remain strong for a long time. In this case the analysis has assisted us in attributing the correct weights to the components of a blend.

V II.

All classified observable lines come from metastable levels, and all behave accordingly. But the unclassified line λ4065.09 which, according to Dunham, contributes largely to λ4065.78 in α Persei cannot be the principal contributor in Nova Herculis, unless the V II line originates from a normal level. In the Nova this strong line rapidly disappears,

as though its lower level is normal. The Fe I line 4065.390 ($z^5F_1^o - f^5D_0$), although originating from a normal level, is not strong enough to account for the intensity in the Nova. The origin of the line in the Nova is doubtful.

Zr II.

All observable lines originate from metastable levels and all behave accordingly.

Nn I.

Only the ultimate lines are observed and these become gradually fainter. But they persist long enough to show that they are insensitive to β.

Ni II.

The few observable lines originate from metastable levels, and their behavior agrees with this conclusion.

Y II.

All observable lines originate from metastable levels and all behave accordingly.

Mg I.

The *b*-group is outside the range of our plates, and some of the components are seriously blended, according to McLaughlin. λ4703 ($3^1P_1^o - 5^1D_2$) disappears rapidly, as it should. The Mg I triplet λλ3829, 3832, 3838 which was discussed on p. 219 is particularly important. McLaughlin has observed it and has been good enough to supply me with information additional to that which he has already published. He has emphasized that λ3838 is the most persistent line of all. This agrees with the theory because its lower level is metastable. Unfortunately, the other two lines were badly blended. The following is quoted from Dr. McLaughlin's letter:

A rough computation from differences of velocity of the absorption systems shows that λ3829 must have been seriously blended at all times after the development of Absorption IV. From January 5 to 12 the center of IV would have been 1.5 *A* to the red of λ3829. There is a gap in our series of plates at that time (January 5.97 to 11.43 G.C.T.). Before the gap λ3829 was plain enough, but afterwards the Absorption IV component of H_η was already so strong and wide that the Mg I line was masked, in spite of the difference of position. Later, IV moved to the violet and its center was never more than 0.6 *A* from λ3829.

The case appeared more hopeful for λ3832. Unfortunately, Absorption III of H_η fell almost precisely on λ3832 early in January. But by January 24 the displacement of III had become 1.0 *A* to the violet of λ3832, and on February 7 it is computed as 1.4 *A* to the violet. On February 20 the new system VII would fall 0.8 *A* to the red of λ3832 (III had disappeared). During March, VII was stronger and only 0.5 *A* from λ3832.

February 7 thus appeared favorable, but at that time VII was just beginning to appear. The plates taken on that date show a line at λ3832 which is much too strong to be Mg I, yet it is not identifiable with either VII or III of H_η. The only interpretation I can suggest is that the observed line is a blend of λ3832, the fading component III of H_η, and the appearing component VII of H_η. Computation shows that the latter two lines should "bracket" λ3832.

MgI Lines in Nova Herculis

Date, G.C.T.	λ3829	λ3832	λ3838	Remarks
1934				
Dec. 28.0	2	3	2.5	
30.5	1.5	2	2	
31.0	1.5	2.5	3	
31.5	1.5	2.5	4	
1935				
Jan. 4.5	2	2.5	3.5	
11.4	bl IV	bl III	3	
12.5	bl IV	bl III	4	
18.5	bl IV	bl III	3.5	
24.4	bl IV	2.5	4	3832 and III resolved
30.4	bl IV	2	4	Resolved?
Feb. 7.4	bl IV	4	3.5	3832 probably blended with III and VII H_η
27.4	bl IV	3 (bl VII?)	3.5	
Mar. 6.3	bl IV	bl VII	3	
22.2	1.5 (IV)	1.5 (VII)	2	Poor plate
23.3	IV	VII	2–	
24.3	2 (IV)	4 (VII)	2	
26.3	IV	VII	2	
27.3	IV	VII	2	

λ3838 was persistent in Nova Aquilæ 1918. In that spectrum λλ3832, 3829 were not blended (unless there are faint unrecognized components of H_η). Estimates of the three lines are given below, but the plates subsequent to June 15 are grainy and are underexposed in the ultraviolet.

MgI Lines in Nova Aquilæ 1918

Date, G.M.T.	3829	3832	3838	Remarks
1918				
June 10.7	3	3	5	
11.7	2.3	3	4.5	3 plates
12.7	1.8	3	4	4 plates
14.7	1	1	2	
15.7	1+	1 ?	1.5	
17.7	1+?	1 ?	1.5	
18.7	?	?	1.5	Poor
22.7	?	?	1.5	
25.7	1 ?	1 ?	1	
July 1.6	2 ?	1	1	
5.6	1	2	1.5	

Dr. McLaughlin concludes that the evidence is inconclusive in both novæ. Nevertheless, the absence of a rapid decline in λ 3832 is probable. Why this should be so is not clear. Perhaps collisions between the three neighboring levels 3^3P^o bring about a rapid redistribution of the atoms, so that the depopulation of level $3^3P_1{}^o$ cannot occur unless the densities are much lower. Since the theory of radiation excitation is reliable and is sufficiently supported by observations, failure to observe a fading of λ 3832 would have no other explanation, and important conclusions regarding the frequency of collisions might be expected. But in view of the uncertainty of the observations it is premature to consider this matter further.

Ca I.

An equally interesting case is that of Ca I. The resonance line λ 4227 is very strong and persists until the end of February or the beginning of March. The multiplet $4^3P^o - 4^3D$ should give valuable results, because one of its lower levels combines with the ground level 4^1S in exactly the same manner as in Mg I. λλ 4455 and 4425 have metastable lower levels, while λ 4435 has a normal level. The line λ 4455 is strong

Multiplet $4^3P^o - 4^3D$ (Ca I)

Lines	I.Q. No.	Lab. Int.	E. P.	Remarks
4425.446	0–1	50	1.9	Unblended, but faint
4434.969	1–2	60	1.9	Bl. Fe I 4435.15 ($a^5D_2 - Z^7F_1{}^o$) of
4435.690	1–1	40	1.9	lab. int. 2 (probably negligible)
4454.795	2–3	80	1.9	Probably essentially unblended
4455.901	2–2	40	1.9	Unblended, but faint
4456.629	2–1	10	1.9	Too faint

and persists a long time, in exact analogy with λ 3838 of Mg I. But λ 4435–6, in which the components of Ca I are not separated, is as strong as λ 4455 on December 28, but is already much weaker on January 1, 2 and 3, and completely disappears thereafter. The relatively great initial strength of λ 4435–6 is due to the blending of the two Ca I components. I have very definitely the impression that the fading of λ 4435–6 is real, in which case it must be caused by the predicted sensitivity to β. But the unfavorable evidence of Mg I rather suggests that we should regard the case as uncertain. The term values of the three Ca I levels are: $4^3P_0{}^o$, 15,157.90 cm^{-1}; $4^3P_1{}^o$, 15,210.06 cm^{-1}; $4^3P_2{}^o$, 15,315.92 cm^{-1}. In Mg I the corresponding levels are: $3^3P_0{}^o$, 21,850.34 cm^{-1}; $3^3P_1{}^o$, 21,870.42 cm^{-1}; $3^3P_2{}^o$, 21,911.14 cm^{-1}. Although the Mg I levels are closer together this difference is hardly sufficient to

produce a large difference in the effects of collisions (which are more effective when the levels are close together).

It is disappointing that we cannot produce a conclusive result. But we can at least point out the importance of securing adequate data for Mg I and Ca I when we again have the good fortune to observe a "slow" nova.

Sr II.

The ultimate lines λλ 4078 and 4216 are strong and persist until the middle of February.

Sc II.

All strong lines have metastable lower levels and behave accordingly.

C I.

The lower levels of the strong lines λλ 4762, 4771, 4776 are normal (multiplet $3s^3P^o - 4p^3P$). These lines change little in intensity until January 5. They rapidly decrease after that date and only a faint suspicion of λ 4776 is present on January 18. The fading is slower than that of Mg II 4481, Fe I 4260, etc. We should have expected that the strong transition from level 3^3P^o to the ground level, by means of the lines at λ 1657, would rapidly depopulate the level, as β decreases. There is no simple explanation for this anomaly, although it must be emphasized that the C I lines fade more rapidly than do lines having metastable lower levels, like Ti II, Fe II, Sc II, etc. We shall return to this question in our discussion of P Cygni (p. 242).

4. In discussing the fading of the lines of Absorption II in Nova Herculis, McLaughlin [19] gave the following lines as fading rapidly:

Line	Identification	
λ4015	Sc II λ4014.532	
4384	Blend	
3947	O I	λλ4947.33; 3947.51; 3947.61
4368	O I	λ4368.30
4481	Mg II	
3983	Y II	

The fading of the O I lines is understandable because these lines have normal lower levels. We have already dealt with Mg II 4481. Sc II 4015 has a metastable lower level. Our plates show a slow gradual decrease from December 28 until January 11 or later, which is quite unlike the rapid decrease of lines having normal levels. But this decrease is, nevertheless, out of harmony with the increase, in the early stages, of the intensity of λ 4247 of Sc II. Probably the Sc II line 4015 is too weak to account fully for the line in the Nova. The blend is not known.

Y II 3983 has also a metastable lower level and should persist as do the other lines. Our plates suggest that this line is very persistent, lasting until the end of January. But in the latter stages it is superposed over the broadened wing of the Ca II — H_ϵ emission. This may give it a fainter appearance than it deserves.

The persistent lines given by McLaughlin in the order Fe II, Sr II, Fe I, Sc II, V II, Ti II, Mg I 3838 are fully accounted for in our theory, but the order of persistence remains unexplained for lack of knowledge of α and γ and of the relative abundances.

I have made no attempt to study other novæ. The material is not sufficient except for Nova Pictoris. H. Spencer Jones' study of the latter [26] fails to bring a conclusive answer. Although Fe I was strong on many of his plates, there is no clear indication of a decrease in such lines as Fe I 4260, as compared to Fe I 4272. But this does not exclude the possibility that effects similar to those observed in Nova Herculis may have been present. A study of the effect in several novæ should be useful to distinguish between the expanding shell hypothesis and the continuous outflow hypothesis.

It is probable that at the time of maximum on December 23, 1934, the photospheric radius was greatly expanded, and Grotrian [27] gives

$$r_{p\ \max} = 100R_\odot = 7 \times 10^{12} \text{ cm.}$$

The radial velocity of Absorption II was, on the average,

$$v = 350 \text{ km/sec.}$$

The maximum distance to which Absorption II had expanded in an interval of 10 days is

$$r_{\text{shell}} = 5r_{p\ \max}.$$

Hence from formula (4)

$$\beta = 0.01. \tag{32}$$

We had estimated from the lines of Mg II, Fe I, etc. that β was somewhere between 0.1 and 0.01, by about January 5. If we remember that the outflow from the photosphere probably continued over an appreciable length of time, it is obvious that (32) represents a lower limit, and that a value near $\beta = 0.1$ may be more nearly correct. At any rate, the computed value of β agrees in order of magnitude with the value derived from the spectrum. But after the beginning of January the velocity of Absorption II changed but little. Accordingly, by the

[26] *Ann. Cape Obs.*, **10**, Part 9, 1931.
[27] *Zs. f. Ap.*, **13**, 219, 1937.

end of March we have approximately

$$r_{\text{shell}} = 50 r_{p\ \max},$$

and

$$\beta = 10^{-4}. \tag{33}$$

Since, again, the thickness of the shell must be considerable, we must regard (33) as a lower limit, and we should probably have to assume

$$\beta = 10^{-3}. \tag{34}$$

This value is not small enough to account for the redistribution of atoms required to make the forbidden emission lines of Fe I approximately equal in intensity to the permitted emission lines. We should rather expect that at the end of March $\beta = 10^{-6}$.

But we must remember that, according to Grotrian [27] and Beileke,[28] the photospheric radius r_p decreased by the end of March to

$$r_p = 10 R_{\odot}.$$

If we allow for this shrinkage, we find

$$\beta = 10^{-6}, \tag{35}$$

in good agreement with the observations.

I do not lay much stress upon these computations, principally because of the somewhat uncertain state of the observational data. But provisionally we may conclude that the spectrographic results are in harmony with Grotrian's model of the Nova.

IV. Other Stars

In preparing the material for the discussion of Fe I in Nova Herculis I have compared a number of F type stars, notably the supergiant ϵ Aurigæ (normal spectrum), the giant α Persei, and the dwarf 41 Cygni. There are no appreciable differences between lines of Fe I which originate from metastable levels and those which originate from normal levels. In order to study possible very small departures from equilibrium in ϵ Aurigæ, it will be necessary to construct new curves of growth for all three stars. This result is not surprising, since we have no other indication that the bright F 5 component is anything but a normal supergiant. It is quite safe to conclude that β is not much smaller in this star than in the other two.

The infra-red component of ϵ Aurigæ is definitely abnormal. But since we never observe its spectrum without some blending and over-

[28] *Die Sterne*, **17**, 25, 1937.

lapping with the F 5 component, we cannot determine β. There is no evidence from the lines of Fe I, because even the strongest lines originating from normal levels are too weak for observation. Mg II is weak in the infra-red star, but is probably not completely absent. This should give us at least a crude idea of β. Strömgren [29] used $\beta = 2.5 \times 10^{-4}$ in his theoretical discussion of ϵ Aurigæ. From the probable existence of Mg II 4481 in the spectrum I should have expected a larger value, but it is quite possible that the normal intensity of the line (without dilution) would be such that it would fall upon the flat portion of the curve of growth. The spectrum of the infra-red star contains no other lines sensitive to β.

The B 5 component of β Lyræ shows [30] only a faint trace of Mg II 4481 and perhaps of Si II 4128. He I is characteristic; the triplet lines are exceedingly strong while the singlets are weak, with the exception of the lines which originate from the metastable 2^1S level. This is the effect discussed by Struve and Wurm.[5] There are no lines of Si III, O II, N II, C II, all of which originate from normal levels. The absence of Si II 4131 throws some doubt upon the identification of Si II 4128. We conclude that the dilution is appreciable.

The spectrum of 17 Leporis [31] also gives evidence of great dilution. The lines of Fe II, Fe I, Ti II, Zr II, Y II, Cr II, Sc II, Ca I, etc. all originate from metastable levels or from the ground level. But the absence of forbidden emission lines of Fe II and, for that matter, the absence of any conspicuous emission lines of Fe II (faint emission borders have been suspected) shows that β cannot be very small. Under these conditions formula (27) applies approximately, and the transition A_{21} plays no role. Hence we should expect that the intensities of the members of multiplets are normal. We have no accurate measures, but the inspection of the spectrograms shows at least approximately normal ratios within the multiplets, and even between different multiplets. It is, however, possible that the Ti II multiplet $b^2F - x^2D^o$, $\lambda\lambda$ 4164, 4172, is abnormally weak.

In this connection attention is called to the fact that when β is small enough for formula (30) to go into effect, a study of the relative intensities of absorption multiplets of Ti II, Fe II, etc., should throw important light upon the difficult question of relative transition probabilities within the forbidden multiplets. The simple sum rules are no longer to be trusted and the special selection rules for quadrupole radiation may well produce anomalous populations in the lower metastable

[29] *Ap. J.*, 86, 604, 1937.
[30] Pillans, *Ap. J.*, 80, 57, 1934.
[31] Struve, *Ap. J.*, 76, 85, 1932.

levels. But few stars are suitable for such a study. Nova Herculis, in the last stage of Absorption II, and perhaps other stars having forbidden emission lines of Fe II, may be useful in this connection.

The weakness of Sr II in 17 Leporis finds no explanation in terms of dilution.

The spectrum of P Cygni contains many absorption lines having normal lower levels. Among these are Si III, O II, Si IV, N III and others. But Mg II 4481 and C II 4267 are abnormally faint and Si II is completely absent. The radial velocities of the absorption lines prove that there is a continuous outflow of gas. There is also definite proof of stratification,[32] elements of low ionization potential occupying the outermost layers. It appears that all cases of strong absorption lines originating from normal levels are produced by atoms of high ionization potential. We infer that in the deeper layers where these lines are formed the dilution is not conspicuous. In the outer layers, where the elements of lower ionization potential are sufficiently abundant, the dilution must be large. This would require a fairly extended atmosphere, but not one of unprecedented dimensions. We have seen that β is proportional to $(R/z)^2$ and that, therefore, an atmospheric height of the order of the star's radius would suffice to produce the observed effects.

The rather obvious correlation between stratification and β suggests that a similar phenomenon may be present in Nova Herculis. It is entirely possible that the shell is stratified and that this explains the relatively slow fading of the lines of C I.[33]

The problem of the peculiar Be stars has already been discussed and there remains little to be added to the work of Struve and Wurm.[5] But the He anomaly in ordinary B stars, supergiants, as well as dwarfs, remains very puzzling. This anomaly was found many years ago and I have tried to explain it as a consequence of dilution, complicated by differences in the curve of growth. But the normal supergiants of class B, for example, β Orionis, 67 Ophiuchi, etc., have strong lines of Mg II 4481, Si II 4128 and 4131. We have found throughout this paper that these lines are very sensitive to β, and, indeed, they are weak or absent in the peculiar Be stars ζ Tauri, ϕ Persei, 48 Libræ, ϵ Capricorni, γ Cassiopeiæ, etc. Hence, I am now inclined to doubt the

[32] Struve, *Ap. J.*, **81**, 66, 1935.

[33] Although the lines of O I are somewhat blended I have the impression that they behave very much like the lines of C I. The early appearance of forbidden [O I] emission has been discussed by Grotrian, but since the lines of the latter have low excitation potentials, while the permitted absorption lines have high excitation potentials, it is possible that they originate from different levels.

correctness of my interpretation of the He anomaly in normal B stars.[34] The only explanation I have to offer is that the anomaly, since verified by E. G. Williams,[35] P. Rudnick [35] and L. Goldnerg,[37] bears a striking resemblance to the dilution effect in tenuous outer shells, where its reality is demonstrated by the sharpness of those lines which originate at high levels. This modification would have been difficult to accept, were it not for the fact that Goldberg [37] has recently found that a complete and satisfactory explanation of the anomaly in normal B stars may be found in the effect upon the curves of growth of the differences between the damping constants of the He triplets and singlets. This does not mean that departures from thermodynamic equilibrium are not present; but they are overshadowed by the curve-of-growth effects discussed by Goldberg. On the whole this leaves the B stars in a much more satisfactory state than before.

I am deeply indebted to Professor Henry Norris Russell, to Dr. Dean B. McLaughlin and to Mr. Leo Goldberg for information concerning various phases of this subject; and to Dr. Theodore Dunham, Jr., for furnishing his unpublished table of wave-lengths in the violet part of the spectrum of α Persei.

[34] *Ap. J.*, **82**, 256, 1935; **88**, 108, 1938; *Pop. Astr.*, **46**, 508, 1938.
[35] *Ap. J.*, **83**, 305, 1936.
[36] *Ibid.*, **83**, 439, 1936.
[37] *Ibid.* In Press.

PLATE I

Lines marked on the margins

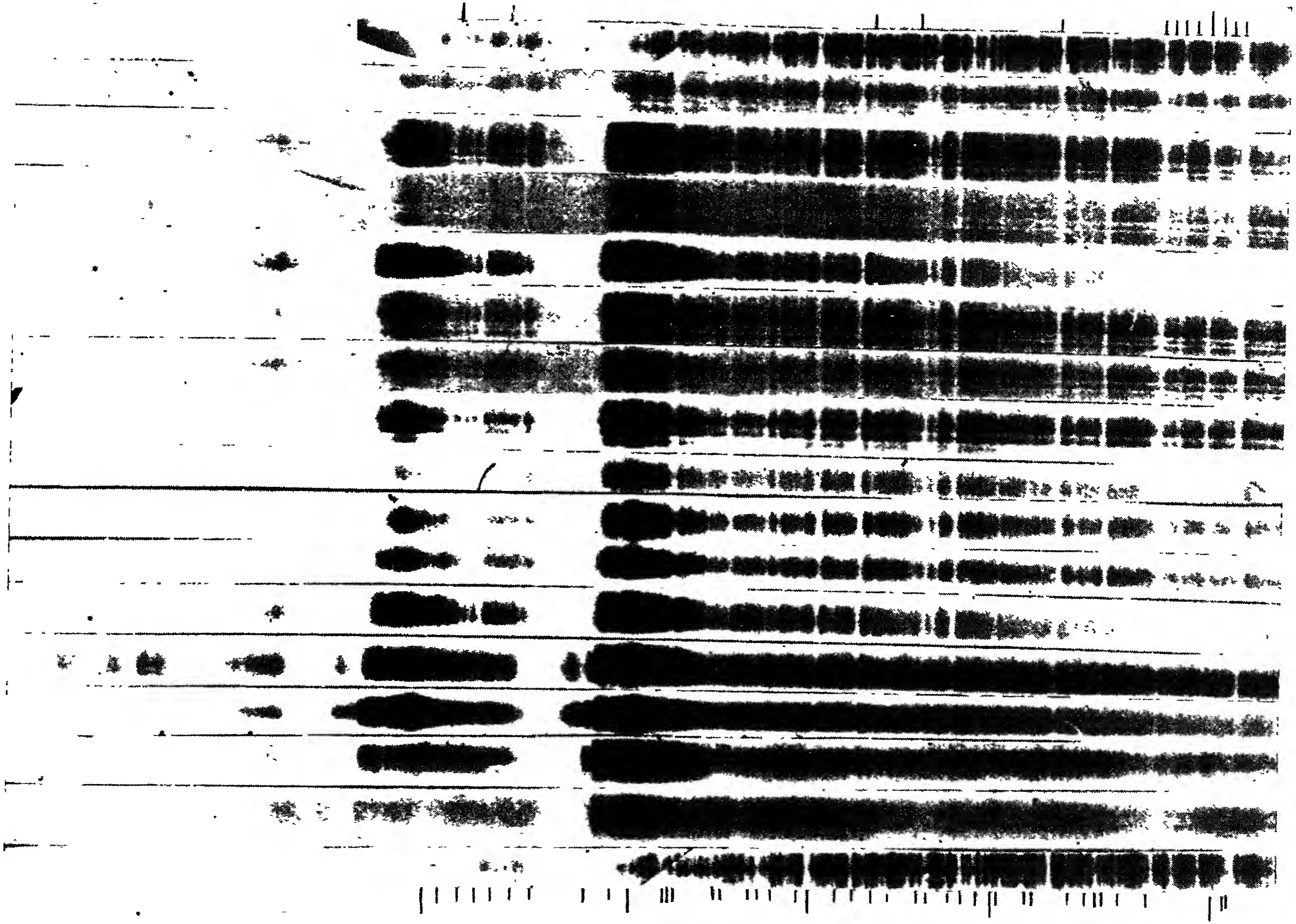
Top

1.	La	II	3949
2.	Fe	I	3956
3.			4015
4.	Fe	I	4022
5.	Fe	I	4045
6.	Fe	I	4063
7.			4065
8.	Ni	II	4067
9.	Fe	I	4069
10.	Fe	I	4072
11.	Ce	II	4074
12.	Ce	II	4076
13.	Sr	II	4078

Bottom

1.	Cr	II	3945
2.	Fe	I	3947–8
3.	Y	II	3950
4.	Fe	I	3953
5.	Fe	I	3955
6.	Zr	II	3958
7.	Ae	I	3962
8.	Ca	II	3969
9.	V	II	3974
10.	Fe	I	3977
11.	Ti	II	3982
12.	Y	II	3983
13.	Fe	I	3984
14.	Ti	I	3990
15.	Zr	II	3991
16.	La	II	3996
17.	Fe	I	3997
18.	Zr	II	3999
19.	Ce	II	4004
20.	Fe	I	4005
21.	Fe	I	4010
22.	Ti	II	4012
23.	Ni	II	4016
24.	Zr	II	4018
25.	V	II	4023
26.	Ti	II	4025
27.	Ti	II	4028
28.	Mn	I	4031
29.	Mn	I	4033
30.	V	II	4036
31.	V	II	4037
32.	Ce	II	4041
33.	La	II	4043
34.	Zr	II	4049
35.	Cr	II	4052
36.	Ti	II	4054
37.	Fe	I	4055
38.	Fe	I	4057
39.	Fe	I	4063
40.	Ce	II	4074
41.	Ce	II	4076
42.	Fe	I	4077

PLATE I



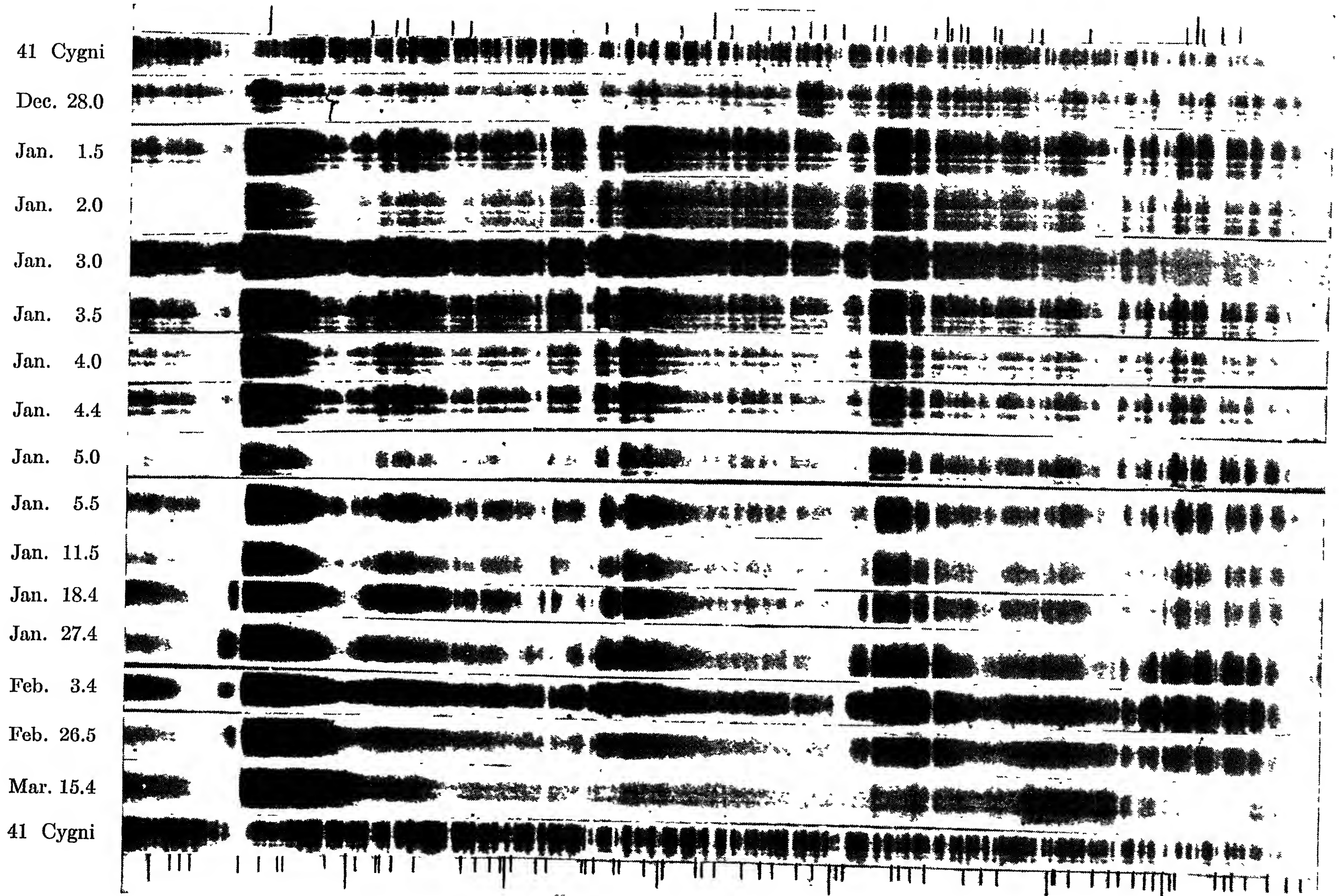
NOVA HERCULIS 1934

PLATE II

Lines marked on the margins

Top				Bottom				Bottom			
1.	Fe	I	4107	1.	Fe	I	4085	46.	Fe	I	4245
2.	Si	II	4128	2.	Fe	II	4089	47.	Cr	I	4254
3.	Si	II	4131	3.	Zr	II	4091	48.	Fe	II	4258
4.	Fe	I	4135	4.	Co	I	4092	49.	Cr	II	4262
5.	Fe	I	4144	5.	H_δ			50.	Fe	II	4273
6.	Fe	I	4148	6.	Fe	I	4106	51.	Cr	I	4275
7.	Fe	I	4176	7.	Cr	II	4111	52.	Cr	II	4276
8.	Fe	I	4182	8.	Fe	I	4113	53.	Fe	II	4278
9.	Fe	I	4191	9.	Fe	I	4119	54.	Cr	II	4284
10.	Fe	I	4198	10.	Fe	II	4123	55.	Ti	II	4288
11.	Fe	I	4202	11.	Y	II	4125	56.	Ti	II	4290
12.	Fe	I	4210	12.	Fe	II	4129	57.	Ti	II	4294
13.	Sr.	II	4216	13.	Eu	II	4130	58.	Fe	II	4297
14.	Fe	I	4219	14.	Fe	I	4132	59.	Fe	I	4299
15.	Fe	I	4222	15.	Ce	II	4138	60.	Ti	II	4300
16.	Ca	I	4227	16.	Cr	II	4146	61.	Ti	II	4302
17.	Fe	II	4233	17.	Zr	II	4149	62.	Fe	II	4303
18.	Fe	I	4236	18.	La	II	4152	63.	Ti	II	4313
19.	Sc	II	4247	19.	Fe	I	4154	64.	Ti	II	4315
20.	Fe	I	4250	20.	Fe	I	4155	65.	Ti	II	4317
21.	Fe	I	4251	21.	Zr	II	4156	66.	Sc	II	4321
22.	Cr	II	4253	22.	Zr	II	4161	67.	Fe	I	4326
23.	Cr	I	4254	23.	Ti	II	4164	68.	Ti	II	4330
24.	Fe	I	4260	24.	Fe	I	4171	69.	La	II	4334
25.	Cr	II	4262	25.	Ti	II	4172				
26.	La	II	4269	26.	Fe	II	4173				
27.	Fe	I	4272	27.	Y	II	4178				
28.	Fe	I	4282	28.	Fe	II	4179				
29.	Sc	II	4306	29.	Ti	II	4184				
30.	Ti	II	4308	30.	Fe	I	4187				
31.	Y	II	4310	31.	Fe	I	4188				
32.	Sc	II	4314	32.	Fe	I	4195				
33.	Ca	I	4319	33.	La	II	4197				
				34.	Fe	II	4199				
				35.	Y	II	4205				
				36.	Fe	I	4207				
				37.	Zr	II	4212				
				38.	Fe	I	4218				
				39.	V	II	4220				
				40.	Fe	I	4224				
				41.	V	II	4225				
				42.	Fe	I	4227				
				43.	Fe	I	4239				
				44.	Fe	I	4240				
				45.	Cr	II	4242				

PLATE II



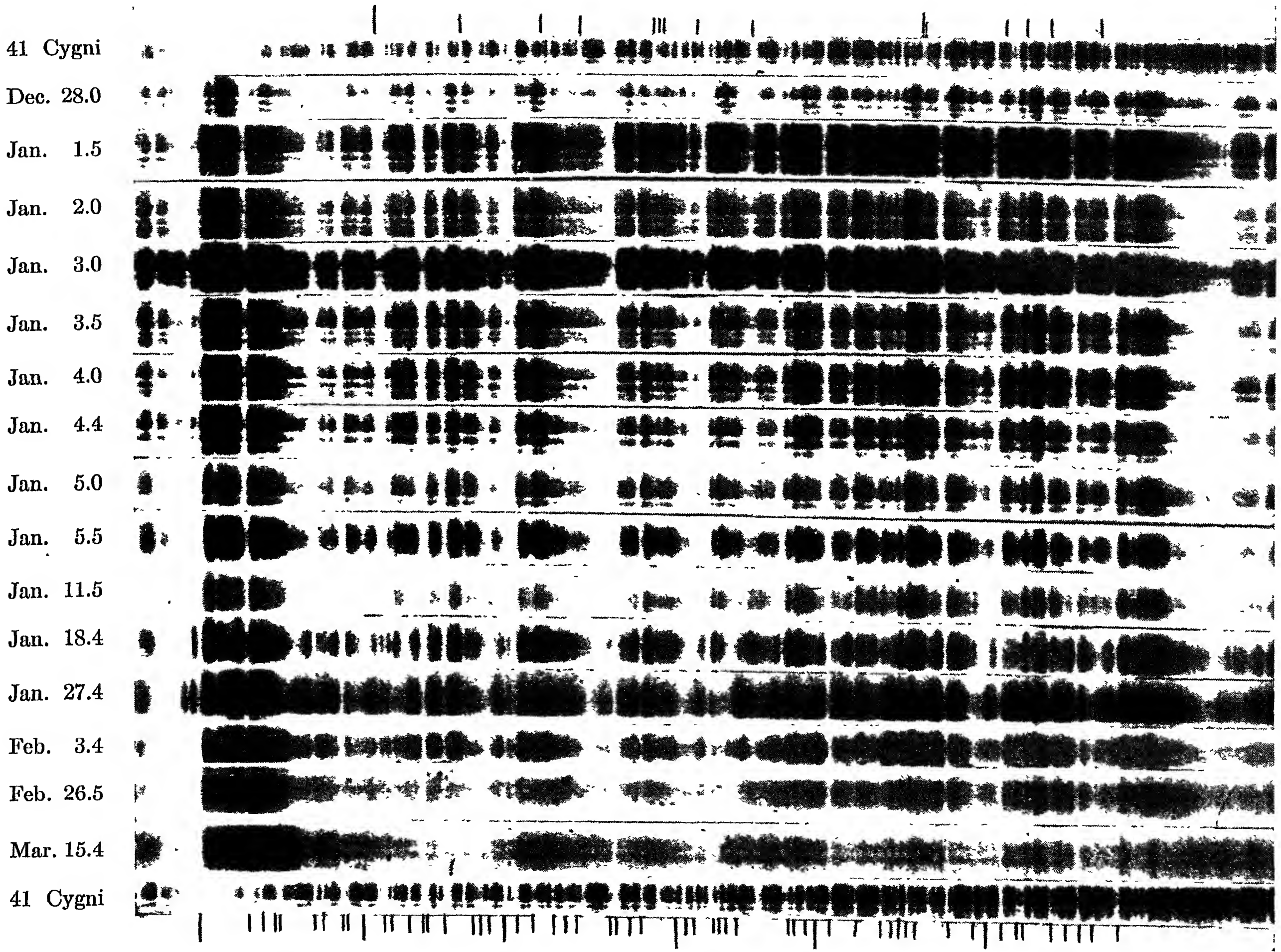
NOVA HERCULES 1934

PLATE III

Lines marked on the margins

Top		Bottom		Bottom	
1.	Fe I 4384	1.	H_γ	46.	Fe II 4523
2.	Fe I 4404	2.	Fe II 4352	47.	Fe I 4525
3.	Ca I 4425	3.	Sc II 4355	48.	Ti II 4534
4.	Ca I 4435-6	4.	Y II 4359	49.	Fe II 4541
5.	Ca I 4455	5.	Zr 4360	50.	Ti II 4544
6.	Ca I 4456	6.	Ti II O I 4368	51.	Ti II 4550
7.	Ti II Zr II 4457	7.	Fe I 4370	52.	La II 4554
8.	Fe I 4466	8.	Sc II 4375	53.	Fe II 4556
9.	Mg II 4481	9.	Fe I 4376	54.	Ti II 4564
10.	Ti II 4529	10.	Zr II 4380	55.	Ti II 4568
11.	Fe I 4529	11.	Fe II 4385	56.	Ti II 4572
12.	Ti II 4552	12.	Ti II 4387	57.	Fe II 4576
13.	Cr II 4558	13.	Ti II 4391	58.	Fe II 4584
14.	4565	14.	Ti II 4394		
15.	Ti II 4580	15.	Ti II 4395		
		16.	Ti II 4400		
		17.	Ti II 4408		
		18.	Ti II 4409		
		19.	Ti II 4411		
		20.	Fe I 4415		
		21.	Ti II 4418		
		22.	Ti II 4422		
		23.	Fe I 4427		
		24.	La II 4430		
		25.	Fe I 4433		
		26.	Fe I 4442		
		27.	Ti II 4444		
		28.	Fe I 4448		
		29.	Ti II 4450		
		30.	Fe I 4459		
		31.	Fe I 4461		
		32.	Ti II 4464		
		33.	Ti II 4468		
		34.	Ti II 4471		
		35.	Fe II 4473		
		36.	Fe II 4476		
		37.	Ti II 4488		
		38.	Fe II 4491		
		39.	Fe I 4495		
		40.	Zr II 4497		
		41.	Ti II 4501		
		42.	Fe II 4508		
		43.	Fe II 4515		
		44.	⊙ II 4518		
		45.	Fe II 4520		

PLATE III



Nova Herculis 1934

PLATE IV

Lines marked on the margins

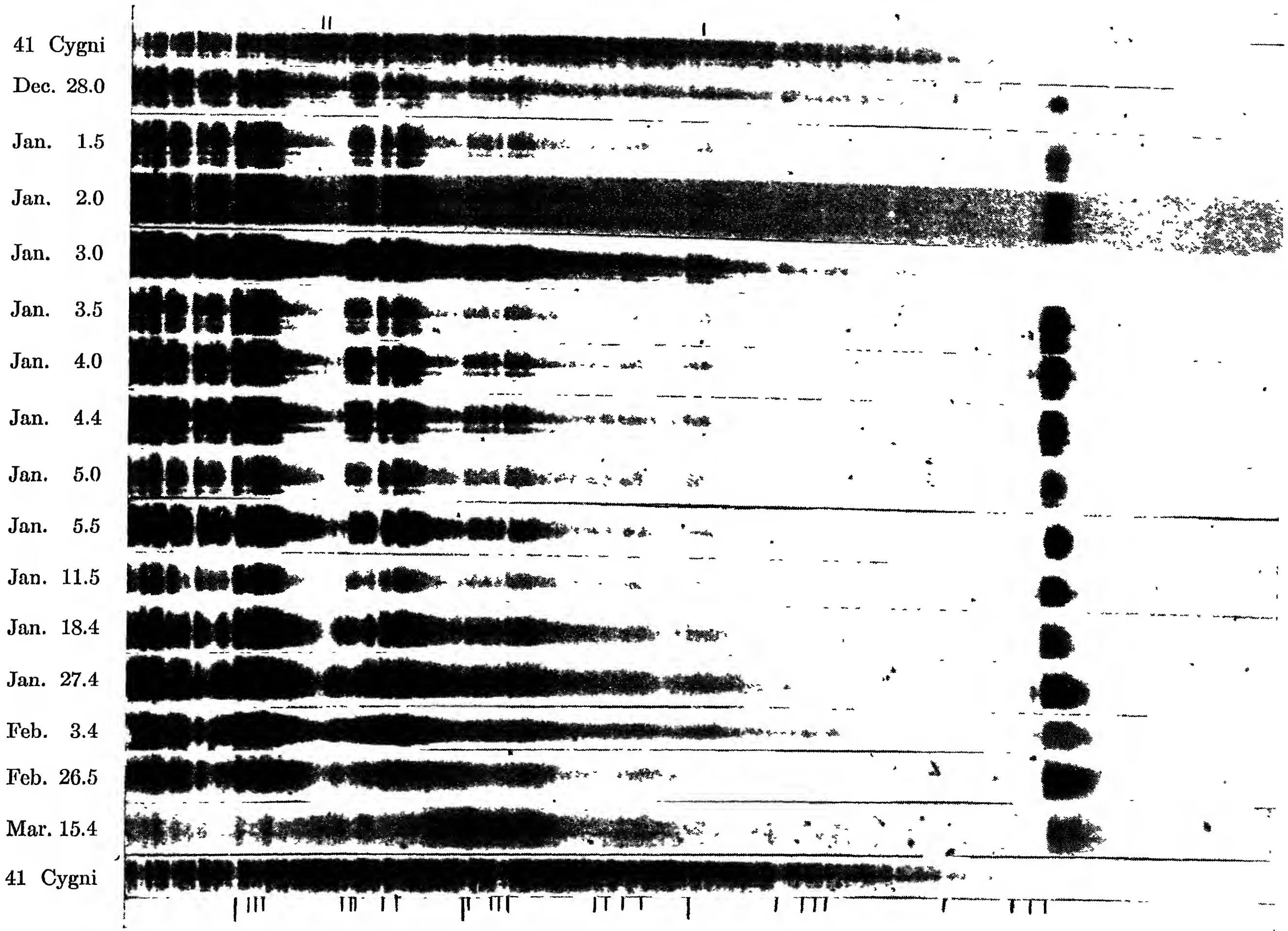
Top

1.	Fe	I	4611
2.	Fe	I	4613
3.	Fe	I	4737

Bottom

1.	Fe	II	4584
2.	Cr	II	4588
3.	Ti	II	4590
4.	Cr	II	4592
5.	Cr	II	4617
6.	Cr	II	4619
7.	Fe	II	4621
8.	Fe	II	4629
9.	Cr	II	4634
10.	Fe	I	4655
11.	Fe	II	4657
12.	Fe	II	4664
13.	Fe	II	4667
14.	Sc	II	4670
15.	Sc	II	4699
16.	Mg	I	4703
17.	Ti	II	4709
18.	Ni	I	4714
19.	Fe	II	4731
20.	C	I	4762
21.	C	I	4771
22.	C	I	4776
23.	Ti	II	4780
24.	Ti	II	4804
25.	Cr	II	4824
26.	Cr	II	4848
27.	Y	II	4855
28.	H_β		

PLATE IV



Nova Herculis 1934

ON THE PHYSICAL CHARACTERISTICS AND ORIGIN OF THE SUPERNOVÆ

FRED L. WHIPPLE

Harvard College Observatory

(*Read February 17, 1939, in Symposium on Progress in Astrophysics*)

THAT the temporary stars comprise two groups with respect to absolute brightness was first recognized by Shapley.[1] The term *nova* has been applied as a generic designation for all such stars, but in recent years the members of the brighter and rarer group have been designated as *supernovæ*. The latter attain on the average a maximum luminosity some ten-thousand times that of the common novæ and fifty-million times that of the Sun. The extreme rarity of supernovæ is illustrated by the fact that none have been observed in our galactic system since the time of Kepler. Discovery had depended upon chance observations of external galaxies until 1936 when a systematic search program in the Coma Virgo cluster was instituted and vigorously pursued by Baade and Zwicky.[2] About one-fourth of the twenty odd supernovæ now recognized have been discovered in this thorough search, and the timely announcements have made possible more complete observational studies.

A discussion of the physical characteristics and possible origin of the supernovæ and their relation to the common novæ depends upon observations of the light variations, frequency, positions in galaxies, and spectra. Only the occurrence and position are known for all the twenty-odd supernovæ certainly identified as such. Seven fairly complete light curves are available but the rising branch of the curve, unfortunately, is not well observed in any case. Spectra have been obtained and their descriptions published for five supernovæ.

Baade [3] and Miss Hoffleit [4] have compiled all the available observations of the magnitudes of the supernovæ and have discussed the light curves. Miss Hoffleit's present discussion eliminates the need for more than a comment here. There can be no doubt of the extremely high absolute luminosities attained by the supernovæ. It is beyond the realm of all reasonable probability that so many novæ could occur ap-

[1] Shapley, *P. A. S. P.*, **29**, 214, 1917.
[2] Zwicky, *P. S. A. P.*, **50**, 215, 1938.
[3] Baade, *Ap. J.*, **88**, 285, 1938.
[4] Miss Hoffleit, These *Proceedings*, this number.

parently as near the nuclei of external galaxies as they do without being physically associated with them. Since the distances of the galaxies are known with fair precision, the absolute magnitudes at maximum light range from -12 to -16 for the various supernovæ. It is quite possible that fainter supernovæ occur to bridge the gap in absolute magnitude between the novæ and supernovæ. The light curves of the two classes are generally similar but the supernovæ appear to fade with less irregularity than the novæ.

The integrated light curves correspond to from 1×10^{47} to 5×10^{48} ergs of energy (E_{vis}), if one assumes that the spectral energy distribution curve is the same as that of the sun. For other energy distributions or temperatures the integrated energy would, in general, be greater. The determination of the total energy involved in the supernova phenomenon constitutes, at the moment, one of the most important problems to be solved before any theory as to the origin of the phenomenon can be seriously upheld. We shall show below the results of various assumptions as to the energy distribution.

The frequency of occurrence of supernovæ has been investigated by Zwicky [5] on the basis of his systematic search program. On the average one supernova occurs every six hundred years in a single galaxy. This frequency is so low and the energy output so great compared to the corresponding quantities for the ordinary novæ that analogies between the two must be employed with great caution. There is no reason to suppose that the origins or methods of energy release are at all similar.

The spatial distribution of the supernovæ with respect to the centers of their respective galaxies has been studied by the author.[6] It was assumed that the supernovæ lay in the principal planes of their galaxies, and corrections were made for the effect of projection. The distribution was found to be somewhat similar to that of the surface luminosity of the galaxies, with a concentration towards the nuclei and in the main spiral arms and luminous knots. Practically all the supernovæ occurred within a distance of 1200 parsecs from the nuclei, *i.e.*, within the "main bodies" of the galaxies as defined by Hubble.[7] This distribution is also similar to that for the common novæ observed by Hubble [8] in the Andromeda Nebula, and suggests that both the ordinary novæ and the supernovæ are produced among the average stars responsible for the principal mass and luminosity of the galaxies. The bright variable stars and other very luminous stars, although there is the

[5] Zwicky, *Ap. J.*, **88**, 529, 1938.
[6] Whipple, *P. N. A. S.*, **25**, 118, 1939.
[7] Hubble, *Realm of the Nebulæ*, p. 178, 1936.

possibility of observational selection, appear more frequently in the outer regions.[8]

The spectra of the various supernovæ are similar in general character. Mrs. Payne-Gaposchkin [9] has suggested that they consist of emission bands widened by the Doppler effect of an expanding atmosphere. Her discussion was based upon the visual observations of S Andromedæ and the Harvard spectrum of Z Centauri. She also suggested that the wide band near $\lambda 4640$ observed by Johnson [10] in the spectrum of Z Centauri might be of the same origin as in novæ spectra, thus indicating a general similarity between novæ and supernovæ with regard to spectra and spectral changes. Subsequent observations and interpretations of supernovæ spectra confirm (with still a remote possibility of change) the existence of wide emission bands but are not as yet conclusive concerning the exact nature of the emission spectra. One conclusion is certain: there is no strong evidence of the hydrogen spectrum, so conspicuous for the common novæ.

Discussions of the spectra of supernovæ have been presented by Ritchey,[11] Humason,[12] Baade,[13] Popper,[14] Strohmeier [15] and extensively by R. Minkowski.[16] No identifications of any of the wide bands common to all of the spectra have been agreed upon by all these investigators. R. Minkowski, whose observations and discussions are the most complete, concludes that the spectrum of any supernova changes considerably in structure with the time in such a way that the features in the photographic region of the spectrum shift progressively towards the red (approximately 100 A in a few months) while the features of the visual region change in character but do not shift in position with the time. His reproduced microphotometer tracings suggest that in the early stages the spectra are nearly continuous and are relatively more intense in the visual regions than the spectra of A-type stars. With the passage of time the visual intensity generally fades with respect to the photographic intensity and the emission bands become more conspicuous.

The spectra of all the supernovæ are remarkably similar with respect to the apparent widths of the emission bands. If these widths are interpreted as representing velocities of expansion of the atmosphere, the

[8] Hubble, *Ap. J.*, **69**, 103, 1929.
[9] Mrs. Payne-Gaposchkin, *Ap. J.*, **83**, 245, 1936; *ibid.*, p. 173, 1936.
[10] Johnson, H. B. 902, 1936.
[11] Ritchey, *P. A. S. P.*, **29**, 211, 1917.
[12] Humason, *P. A. S. P.*, **48**, 110, 1936.
[13] Baade, *P. A. S. P.*, **48**, 226, 1936.
[14] Popper, *P. A. S. P.*, **49**, 283, 1937.
[15] Strohmeier, *Zs. f. Ap.*, **14**, 227, 1937.
[16] R. Minkowski, *Contr. Mt. Wilson Obs.*, No. 602, 1939.

velocities (from half-widths of bands) are of the order of 5000 to 7000 km/sec, the uncertainty of measurement being of the same order as the range in velocity between various objects.

The velocities of expansion for five supernovæ, identified by the galaxies in which they appeared, are given in the second column of Table I. The values were adopted from the various sources indicated

TABLE I

VELOCITIES OF EXPANSION FOR SUPERNOVÆ

NGC	Vel Expansion (km/sec)	Abs. Mag. at Max.	E_{vis} (ergs)
		m	
1003	6000–7000[16, 17]	−14.0	6×10^{47}
4273	6000[12]–7000[18]	−12.4	2×10^{47}
4303	5800[12]	−11.9	
I4182	4500[14]–6000[16]	−16.2	4×10^{48}
5253	5000[9]	−17 ?	

by the references and must be considered as approximate because of the observational difficulty in determining precise band widths. The absolute magnitudes at maximum light (column 3) and the integrated luminosities (E_{vis}, column 4), derived on the basis of a solar energy distribution, were taken from Miss Hoffleit's [4] results.

We note that there is no evidence for a correlation between the absolute luminosities of the supernovæ and their velocities of expansion, as would be expected from Zwicky's [18] theoretical relations derived for novæ and supernovæ. On the basis of certain assumptions concerning the expansion of material in a stellar outburst, Zwicky concluded that the absolute magnitude at maximum light should vary as the negative logarithm of the fifth power of the expansional velocity, and that E_{vis} should vary directly as the third power of the velocity. From Table I we see that the first relation would require SN IC 4182 to show seven times the expansional velocity of SN NGC 4303, and that the second relation would require SN IC 4182 to show about three times the velocity of SN NGC 4273. Even when we take into account the observational uncertainties in both the velocities and the absolute magnitudes of the supernovæ, it is still clear that the theoretical relationships are not obeyed.

For a working hypothesis, however, the interpretation of band widths as expansion seems to be well justified at the present time. Minkowski considers other possibilities but finds that none is free from damaging

[17] Humason, *H. A. C.*, No. 429, 1937.
[18] Zwicky, *P. N. A. S.*, 22, 457, 1936.

contradictions. The expansion hypothesis is consistent with Humason's and Strohmeier's identifications of certain lines in the spectra as those of hydrogen, helium, nitrogen and oxygen. Although these identifications are subject to question (particularly those for hydrogen, since the Pickering series of ionized helium will account for the possible faint lines observed near the positions of the Balmer lines), they must be considered as presenting a plausible though incomplete explanation of the difficult spectra. A similarity in character between the spectra of supernovæ a few weeks after maximum and the spectra of Wolf-Rayet stars [19] is strongly evidenced by weakness or absence of forbidden lines, the probable presence of helium, and the great widths of the lines resulting from expansion or expulsion of matter. Hydrogen lines are sometimes missing in the spectra of Wolf-Rayet stars. The analogy between the supernovæ and Wolf-Rayet stars is strengthened by the smoothness of the light curves and the constancy of band width of the supernovæ in comparison with those of novæ. The ejection process appears to be uniform, as in the case of the Wolf-Rayet stars, rather than spasmodic and chaotic as in the case of the common novæ, where large irregular masses of gas are known to be ejected.[20] These masses of gas, large in comparison with the amount of matter ejected continuously from a common nova's surface, become in time the chief source of its photographic and visual radiation, which arises from atomic processes induced by the high-temperature "diluted" radiation of the central star. Because of the low density of the matter and the dilution of the radiation, the "forbidden" lines predominate. For the Wolf-Rayet stars, and, we presume, for the supernovæ, prodigality in the ejection of matter does not occur. The spectral lines are emitted near the surface of an effective photosphere or in an extended atmosphere where matter is being continuously ejected. The radiation is not greatly diluted, and the permitted lines that are produced in matter of moderate density near the effective photosphere are much stronger than the forbidden lines that are produced in the matter that has been ejected beyond this region.

This interpretation of the supernova spectra, although very tentative, is, in the author's opinion, generally consistent with observation and physical theory. The observed changes in the spectra with time, from a complex of overlapping lines or continuum to the development of strong individual lines, are to be expected from a general rise in the state of excitation, such as is observed but not completely explained for

[19] *Trans. Int. Ast. Union*, 5, 178, 1936.
[20] Whipple and Mrs. Payne-Gaposchkin, *P. N. A. S.*, 22, 195, 1936.

the novæ. Minkowski's observation of a progressive red shift of a few features in the photographic spectrum without a concomitant shift in the visual region seems inexplicable except as a chance result of changes in the relative intensities of involved emission lines.

The absence or weakness of the hydrogen spectrum is of considerable theoretical importance. It indicates that the state of excitation is very high or else that hydrogen is not abundant. The first alternative is consistent with the occurrence of very high temperatures for supernovæ as suggested by Baade and Zwicky.[21, 22] The alternative of a low absolute abundance of hydrogen has implications that are worthy of consideration. The outer atmosphere of an average star, the Sun for example, consists, according to present theories, almost entirely of hydrogen with some helium and only a trace of the metals and other elements.[23] The main body of an average star, however, consists, according to theories of stellar interiors,[24] of only a moderate fraction of hydrogen with comparable amounts of helium and of the other elements together. In an ordinary nova only a very small fraction of the stellar mass is blown away.[25] This mass presumably comes from the outer layers of the star and is, therefore, mostly hydrogen, in keeping with the great strength of the hydrogen lines in most novæ. In the supernovæ, as we shall show below, much more matter is probably ejected, necessarily from the lower levels of the star. The hydrogen abundance as evinced by the spectra should, therefore, be much less for supernovæ than for common novæ, while the helium abundance may be greater.

The determination of the total energy radiated by a supernova depends upon the distribution of the energy in the entire spectrum. This distribution depends again upon the physical and radiative structures of the surface or volume from which the radiation is emitted. We cannot determine the energy distribution until the spectra are positively interpreted, and it is difficult to interpret the spectra without knowing at least the general order of the excitation (or, roughly, the temperature). This impasse can at present be avoided only by the application of somewhat hazardous assumptions. For S Andromedæ, Baade and Zwicky[21] have calculated that if the temperature were 200,000 degrees the total energy of radiation would be 3×10^{51} ergs. In other calculations they[26] have assumed higher temperatures and obtained total energies of the

[21] Baade and Zwicky, *P. N. A. S.*, **20**, 254, 1934.
[22] Zwicky, *Ap. J.*, **88**, 522, 1938.
[23] Menzel, *Pop. Ast.*, **47**, No. 3, 1939.
[24] B. Strömgren, *Ap. J.*, **87**, 520, 1938.
[25] Whipple and Mrs. Payne-Gaposchkin, *H. C.* 413, 1936.
[26] Baade and Zwicky, *Phys. Rev.*, **45**, 138, 1934.

order of 4×10^{53} ergs. Below we shall supplement their calculations for assumed temperatures at the lower end of the temperature scale.

In an expanding atmosphere or shell of gas the composition of the escaping stream of continuous radiation probably does not closely approximate that of a black-body at any temperature,[25] and the presence of emission lines complicates the spectra even more. When the emission lines and continuum can be directly observed, Zanstra [27] has shown, by use of certain assumptions, how the temperature and total energy can be determined. Zwicky [28] has applied methods of this type and found relatively small lower limits for mass and radiation losses for supernovæ. A more reliable method at present is probably to assume that the radiation is emitted in an extended atmosphere and to apply the theory developed by Kosirev [29] and Chandrasekhar [30] for the continuous ejection of matter in a nova. The theory will be here applied as it was for Nova Herculis by Mrs. Payne-Gaposchkin and the author.[25]

The density of matter is assumed to vary inversely as the square of the distance, r, from the center of the supernova and directly as the total intensity of radiation. The velocity of ejection is considered as constant. The mass absorption coefficient is assumed to vary as T^{-n} where $n = 11/3$. The case $n = 5$, used also for Nova Herculis, gives generally a smaller total energy than $n = 11/3$. The numerical value of the absorption coefficient is here calculated as it was for Nova Herculis, although a minor correction should be applied if helium is to predominate instead of hydrogen. The bolometric correction for an extended atmosphere is found to be smaller at high temperatures than for black-body radiation comparable in amount, because the energy-distribution curve is flatter for the extended atmosphere.

Calculations were carried out for the supernova in IC 4182 because it attained the greatest absolute luminosity of any supernova and also the greatest integrated brightness. We assume that at maximum light its absolute photographic magnitude was -16.2 (color-index zero) and that its integrated light corresponds to 5×10^6 seconds at the maximum rate of emission, from Miss Hoffleit's data. The velocity of ejection is taken at 6000 km/sec.

In Table II calculations are shown for six assumed values of the effective temperature from 9,500° K to 100,000° K. The corresponding values of the temperature, T_1, at optical depth unity are given in the second line, and the radial distance, r_1, of this layer from the center, in

[27] Zanstra, *Ap. J.*, 65, 50, 1927.
[28] Zwicky, *P. N. A. S.*, 22, 557, 1936.
[29] Kosirev, *M. N.*, 94, 430, 1934.
[30] Chandrasekhar, *M. N.*, 94, 444, 1934.

TABLE II

CALCULATED ENERGY LOSSES FOR THE SUPERNOVA IN IC 4182

T effective ($\times 10^{-3}$ deg. K)	9.5	20	40	60	80	100
T_1 ($\times 10^{-3}$ deg. K)	8.3	17	35	52	70	87
r_1 (a.u.)	24	7.6	3.0	1.9	1.8	1.8
Bolometric Correction	$0^m.59$	$1^m.26$	$2^m.26$	$3^m.1$	$4^m.0$	$5^m.1$
log (Mass Ejected, gm. sec^{-1})	26.2	25.9	25.7	25.6	25.7	25.9
log (Total Mass Ejected, gm)	32.9	32.6	32.4	32.3	32.4	32.6
log (Radiation, ergs sec^{-1})	42.2	42.5	42.9	43.2	43.6	44.0
log (Total Radiation, ergs)	48.9	49.2	49.6	49.9	50.3	50.7
log (Total Mass Energy, ergs)	50.2	49.8	49.6	49.6	49.7	49.8
log (Total Energy, ergs)	50.2	49.9	49.9	50.1	50.4	50.8

the third line. These quantities, and also the mass and radiation lost per second, refer to the time of maximum light. The total mass energy is the kinetic energy of the total mass lost with a velocity of 6000 km/sec. The neglect of the potential energy is not important unless ejection occurs at a surface of radius considerably less than that of the Sun, or unless the star is much more massive than the Sun. The total energy is the sum of the total radiation and the total mass energies.

We note that the calculated value of the total energy lost does not change greatly over the ten-fold range in temperature. The radiation energy increases continuously over this range and continues to do so for higher temperatures because of the increasing bolometric correction resulting from concentration of light in the far ultra-violet. The kinetic energy carried by the ejected matter decreases with temperature for the first part of the table because of the great dimensions of the regions of high optical depth at low temperatures. These opposite tendencies of the radiation and mass energies with temperature keep the total energy near 10^{50} ergs until an effective temperature of 80,000° is reached. Above this value the total energy increases rapidly.

An estimate of the minimum effective temperature for the supernovæ can be determined from the values of r_1 and the time of rise to maximum light by a method similar to those used by Gaposchkin [31] and by Baade and Zwicky.[21] The outermost edge of the ejected material must have traveled at least twice the distance r_1 for the equations to apply at maximum light. For a velocity of 6000 km/sec, or 3.3 a. u. per day, the time of rise would have to be approximately two weeks for the temperature to be as low as 9500°. About five days would result in a minimum effective temperature of 20,000°. Although the total time of rise is not known for any supernova, this method can be applied for

[31] S. Gaposchkin, *H. B.* 899, 1935.

an order-of-magnitude determination of the minimum temperature by means of the few observations of the rate of rise. From the light curves compiled by Miss Hoffleit we find the following rates of rise: S And, $>3^m$? in 1^d; Z Cent, $>5^m$ in 20^d; NGC 2535, $1^m.7$ in 8^d; NGC 4636, $1^m.5$ in 1^d; and IC 4719, $1^m.4$ in 4^d. The total rise in magnitude is probably of the order of 20^m but the observations are necessarily near maximum light. We should expect the very first part of the light curve to be much steeper. The observations above would then indicate that the total time of rise is almost certainly less than ten days and probably longer than three days. On this basis the minimum effective temperature at maximum light would lie between 12,000° and 25,000° K.

It is important to note that almost identical values for the minimal temperatures are obtained if the calculations are based on the simpler theory of an expanding envelope or photosphere radiating as a black body, the rate of expansion being taken as 6000 km/sec.

If the spectrum of a supernova at maximum light consists entirely of overlapping bright lines without an appreciable continuum, the minimal temperatures obtained above lose most of their significance. The general state of excitation can then only be obtained when the outstanding lines are certainly identified. In either case, however, Minkowski's [16] observation of a relatively faint ultra-violet spectrum for IC 4182 to λ3300 is evidence that the effective temperature is not particularly high, unless, of course, some process of selective scattering is in operation.

The effective radius for a supernova at maximum light is seen to be rather large (Table II) in the temperature range assumed. So long as any appreciable continuum from approximately black-body radiation is supposed to exist in the spectrum at maximum light, the calculated radius cannot be reduced below that of the Sun (0.005 a. u.) by the adoption of effective temperatures less than many millions of degrees. This limit arises from the fact that at high temperatures the intensity of radiation in the visual and photographic spectra is given by the Rayleigh-Jeans law in which the intensity increases as the first power of the temperature. Thus the effective radius varies slowly as the inverse square-root of the temperature for high values. Under the limitations of the assumptions (appreciable continuum and approximately black-body radiation), then, we can hardly expect radii smaller than that of the Sun and therefore can expect no important gravitation effects to be manifest in the spectra of supernovæ.

In quitting the discussion of the related subjects of energy output, effective temperature and effective radius of supernovæ, we especially

emphasize that the calculations are all based upon the observed data for the most luminous supernova, the one in IC 4182. The calculated values of energy and radius at an assumed temperature will be reduced in amount for supernovæ of smaller absolute luminosity, and the limiting values of the temperature will be reduced.

In view of the uncertainties concerning the physical and radiative structures of supernovæ and the changes in these structures with time, it is obvious that theories of origin can as yet be no more than hypotheses. A working hypothesis for the origin of supernovæ must lead to an astronomically explosive release of from 10^{48} to 10^{53} or more ergs of energy. Since we know of no energy sources independent of matter, a reasonable hypothesis will automatically associate the energy with matter, and will, therefore, by physical laws, lead to an explanation of the spectra that is satisfactory from our present inchoate viewpoint. Many details must eventually be explained but at the moment it is necessary to find a sufficiently great source of energy that may be released quickly.

The only source of energy so far suggested for the supernova phenomenon is the potential energy in or between stars. Subatomic sources, from the transformation of mass into energy, may also be involved, but it is necessary to show in detail how they could be released. Potential energy might be released by the collapse of a star or by the collision of two stars. Milne [32] and Chandrasekhar [33] have suggested that a star which, by radiative processes, arrives at a critical mass and luminosity might become unstable and collapse into a degenerate state. In so doing it would release potential energy and become a nova or supernova. The energy involved in such a collapse would be of the right order. The definite possibility of this process is shown by the theory of degenerate matter in stellar interiors as developed by the above two authors, and therefore deserves careful consideration. It is hoped that more theoretical work will be done to show the probability of such stellar instability and the details of the processes that should accompany it.

The amount of energy that could be theoretically released from the collapse of a star is limited only by the compressibility of the matter. Higher densities (up to 10^{14} gm. cm^{-3}) than those of degenerate matter might be attained by matter that has been transformed into neutrons. It has been suggested by Baade and Zwicky [34] that a supernova results from the collapse of a star or part of a star to a neutron core. Since

[32] Milne, *M. N.*, 91, 4, 1930.
[33] Chandrasekhar, *M. N.*, 95, 226, 1935.
[34] Baade and Zwicky, *P. N. A. S.*, 20, 259, 1934.

energy would be required to effect the transformation of ordinary matter to neutrons, part of the potential energy would not be available for the observed processes. The remaining energy would, however, still be greater than that lost through the collapse to a degenerate state because of the extremely small dimensions of the neutron stars. Zwicky [22] has mentioned resultant radii as small as 74 km for a star of one solar mass! The neutron-core theory provides an abundance of energy for the supernovæ but is subject to question until there is proof that a neutron core can actually exist and until the theory of stellar interiors demonstrates the possibility of an instability that might produce the collapse. The adjunctive suggestion by Baade and Zwicky that cosmic rays are produced by supernovæ has not as yet been demonstrated.*

The possibility that supernovæ may be produced by the collision of two stars has recently been discussed by the author.[6] Estimates of the space densities of stars in galaxies indicate that collisions should occur from one to thirty per cent as frequently as Zwicky has calculated that supernovæ occur. Since the calculated frequency of collisions has increased by a factor of 10^8 within the past two decades as the result of observational and theoretical advances in our knowledge of the galaxies, the remaining discrepancy of 10^2 or less does not appear to be a serious objection to the collision hypothesis. A more serious objection lies in the fact that collisions of stars comparable to the Sun in mass would provide a quantity of energy near the lower limit of the probable energy expended by a supernovæ, *i.e.*, of the order of 10^{48} ergs except in cases of very central encounters. It is true, however, that very high pressures and temperatures would exist for a short time in large volumes of the two stars while they were in contact. These high temperatures would accelerate nuclear transmutation processes of the type proposed by Bethe [35] for the generation of energy in main-sequence stars. The carbon-nitrogen cycle operates proportionally to the seventeenth power of the temperature; hence temperatures only a few times that at the center of the Sun might provide sufficient energy. Such temperatures (10^{8+} degrees) appear to be possible in stellar collisions. It is also true that encounters would greatly disturb the equilibria of stars and thus might set off processes of collapse.

The general observations of supernovæ appear to be of little assistance in discriminating between the theories of their origin. The calculated values of the total energy vary over a range of 10^5, but because very large values are possible they tend slightly to favor the collapse

* From a conversation with Professor Vallarta.

[35] Bethe, *Phys. Rev.*, 55, 434, 1939.

theories. Since the shapes of the light curves have not been calculated on the basis of any hyopthesis of origin, the observed light curves do not discriminate. The probability of a stellar collapse has not been calculated and cannot be calculated until much more is known about stellar evolution and the frequency distributions of stars with respect to mass and spectral type. Thus the observed frequency of supernovæ supports the collision hypothesis apathetically. A certain knowledge of the state of excitation or temperature of a supernova would probably be of more value as a key to the determination of the total energy than as a direct criterion of the mode of origin. It is quite possible that any quick release of a given quantity of energy in a star's interior would produce a similar observable result regardless of the mode of origin.

The distribution of supernovæ with respect to the nuclei of the galaxies again does not discriminate between the modes of origin. It is approximately the distribution to be expected for the collision hypothesis and for the frequent occurrence of stars of moderate mass.

Even definite evidence that cosmic rays are not produced by supernovæ would not vitiate the hypothesis of collapse to a neutron core. An outer layer of opaque matter might well absorb and transform the cosmic rays to lower frequencies. The existence of such a layer is also indicated by the above considerations of radius and temperature. Even if, contrary to assumption, there is actually no continuum in the spectrum of a supernova at maximum light, the presence of emission lines with widths approximately proportional to the wave-length is still fair evidence for an outer volume of expanding gas.

OBSERVATIONS OF SUPERNOVÆ

DORRIT HOFFLEIT

Harvard College Observatory

(*Read February 17, 1939, in Symposium on Progress in Astrophysics*)

Ordinary novæ, it will be recalled, are those objects that suddenly emerge from the obscurity of apparently commonplace stars, and, within a day or two, become about 20,000 times more luminous than our sun. Supernovæ appear with similar abruptness, but they surpass ordinary novæ in intrinsic brilliance by a factor between one and ten thousand. Mrs. Gaposchkin [1] has taken exception to the term "supernova" on the ground that "nova" signifies "newness," whereas a supernova does not surpass the ordinary nova in this respect. The Swedish astronomer, Lundmark, from the analogy that stars more luminous than giants are called supergiants, was apparently the first to suggest the expression "supernova" in a paper on pre-Tychonic novæ.[2] The term had not been employed by Stratton in the section on Novæ in the Handbuch der Astrophysik (1928), nor do I find it used in any of the references for the years 1928–1932 listed in Helen B. Sawyer's bibliography on supernovæ.[3] The terms giant- and dwarf-novæ had previously appeared, but they were obviously unsatisfactory, since even a faint nova has, at maximum, supergiant luminosity. As the expression "supernova" has now become firmly rooted in the literature, the adoption of a more logical terminology would, as Mrs. Gaposchkin [1] has remarked, only result in confusion. Moreover, in using the word "supernova" we err linguistically only as much as we do in our common usage of the word "novelty" in the sense of unusual; in unusualness a supernova does surpass the ordinary nova. The relative frequency of normal novæ to supernovæ has been roughly estimated at 15,000 to one—normal novæ occurring at the rate of twenty to thirty per year both in the local galaxy and in the Andromeda Nebula, whereas supernovæ, according to a recent estimate by Zwicky, occur one per galaxy every six hundred years.

The mean absolute magnitude of the normal novæ is about −6, with

[1] *Cf. Variable Stars*, p. 269, 1938.
[2] *Lund Circular*, No. 8, 1932.
[3] *J. R. A. S. Can.*, 32, 86, 1938.

a spread of roughly four magnitudes. The average absolute magnitude of the supernovæ is nearly −15, with a similar spread. Very few objects with intermediate magnitudes between -8^M and -12^M have been found. A probable example is the nova of 1860 (T Scorpii) which, if it is a member of the globular cluster Messier 80 in which it appears, has an absolute magnitude of −9; if it is merely a foreground star superimposed on the cluster, its absolute magnitude is normal. Another possible example is the peculiar nova in η Carinæ.[4] The majority of the supernovæ have been discovered in external galaxies too distant for the resolution of stars as faint as normal novæ at maximum. It had therefore been inferred that the absence of novæ with well verified intermediate absolute magnitudes (-8^M to -12^M) was simply the result of observational selection: the intermediate novæ and supernovæ occur too infrequently to have been observed in significant numbers in the local group of galaxies, and the intermediate novæ are still too faint for easy detection in the more distant galaxies. This suggestion, though not yet adequately disproved, does not appear to afford a probable explanation for the gap in the observed frequency distribution between common novæ and supernovæ, since post-maximum observations corresponding to absolute magnitudes between −9 and −12 are available for at least ten verified supernovæ.

Probably the first estimate of the frequency of supernovæ was made by Lundmark in the paper on pre-Tychonic novæ.[5] His analysis of ancient observational records (mainly Chinese) indicated that the frequency of supernovæ might be of the order of one per century for the galactic system. He himself did not give much credence to this determination, as no well attested supernova in our galactic system has appeared since Tycho's nova of 1572—unless possibly Kepler's star of 1604 was also a supernova. It scarcely seems credible that approximately thirty supernovæ should have occurred in the first millennium A.D., whereas at most three (those of 1054, of which the Crab Nebula is believed to be the remains, 1572, and 1604) appeared in the second. Lundmark's estimate is simply of historic interest. It differs by a factor of six from Zwicky's recent estimate for the extra-galactic nebulæ.

H. D. Curtis at the Lick Observatory was the first [6] to suggest the usefulness of a special photographic program for the discovery of novæ in extra-galactic systems. At that time, only existing Lick Observatory plates were examined. Later, in 1919, a similar search was carried out at the Mount Wilson Observatory. It was not until 1928 that external

[4] *Cf.* Bok, *H. R.* 77, p. 73, 1932.
[5] *Op. cit.; cf.* also, *Bergstrand Festskrift*, p. 89, 1938.
[6] *L. O. B.*, 9, 109, 1917.

galaxies were systematically photographed (at first by Hubble, later by Baade with the 10-inch Cooke refractor at Mount Wilson) for the express purpose of discovering supernovæ. Since September 1936 Zwicky at Palomar has successfully employed the 18-inch Schmidt camera in the search for supernovæ; during the first nine months three hundred photographs of the Virgo Cluster and of nearby nebulæ were obtained. They showed between 5,000 and 10,000 nebular images and yielded one supernova [7]—that in NGC 4157.

Before the beginning of the systematic surveys at Mount Wilson and Palomar, fourteen extra-galactic supernovæ had been discovered. A few others had been announced but were subsequently proved to be only plate defects or foreground variable stars belonging to our own galaxy. Since 1936 the list of accepted supernovæ has been increased to twenty-one, four of the new objects having been found by Zwicky.* Reasons may still be found for rejecting one of the twenty-one objects, SS Ursæ Majoris (1909), as a possible foreground variable star. Its light curve is peculiar and its apparent position in the nebula unusual for a supernova (it is nearly ten times as far from the center of the nebula as the average supernova).

Of the twenty-one possible supernovæ, only one was discovered visually—S Andromedæ, the great supernova of 1885—but we also have photographic evidence for its existence. Fig. 1 is a copy of probably the first photograph ever obtained of a supernova. It was taken at the Harvard Observatory on November 3, 1885, two months after the visual discovery of the star. The supernova cannot be distinguished from the nucleus of the nebula; but on the original plate the nucleus is apparently brighter than on any other Harvard plate of the nebula obtained with the same effective exposure. The brightest part of the nucleus is comparable with star *a* on this plate, whereas ordinarily it is comparable only with star *b*. S Andromedæ is situated within the densest part of the nucleus of the nebula (the projected distance is approximately twenty-five parsecs from the center). No other supernova has as yet been discovered closer than about one hundred parsecs to the center of the nucleus. Considering the differences in the apparent concentration of the nucleus of a nebula when photographed with various exposures, we infer that a few supernovæ appearing within the nuclei of nebulæ must have escaped detection.

It is of interest to note the types of galaxies in which supernovæ have been found. Baade [8] has pointed out that, although only thirty-two

[7] Zwicky, *P. A. S. P.*, **49**, 204, 1937.

* Since the presentation of this paper, Zwicky (*P. A. S. P.*, **51**, 36, 1939) has announced four additional supernovæ in faint galaxies, thus bringing the total to twenty-five.

[8] *Ap. J.*, **88**, 285, 1938.

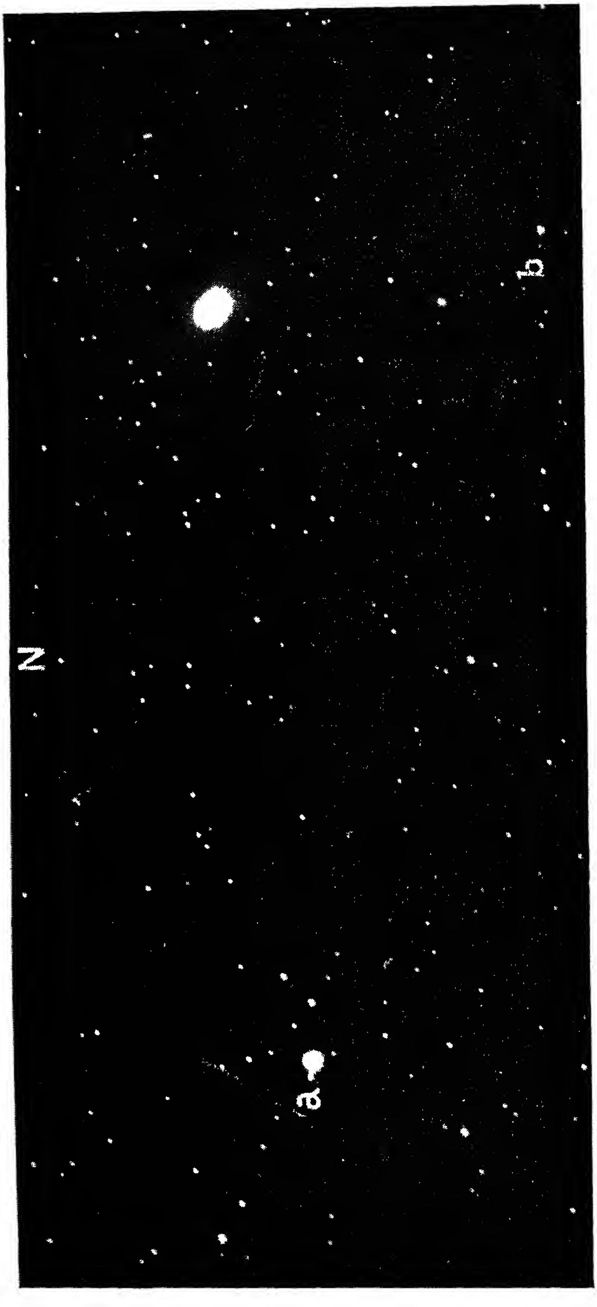


Fig. 1. Andromeda Nebula when the supernova was bright. (Original plate badly scratched.)

per cent of the galaxies brighter than the thirteenth apparent magnitude are late type spirals, seventy-two per cent of the known supernovæ have been found in them.

Zwicky's Seven Characteristics for Supernovæ. In a series of theoretical papers appearing between 1934 and 1937 Baade and Zwicky enumerated possible characteristics for supernovæ; they are summarized by Zwicky in a recent paper [9] of which we give the following outline:

1. Absolute magnitudes of the order -14^M, or comparable with the integrated magnitudes of the nebulæ in which they appear.
2. Energy radiated per year from 10^{48} to 10^{49} ergs for λλ 3,800 to 6,500.
3. Spectrum different from that of any other known stellar object.
4. Frequency, one per galaxy per several centuries.
5. Absolute surface-temperatures at least several hundred thousand degrees.
6. A source for cosmic rays.
7. Cause, the conversion of a normal into a neutron star.

The first, second, and fourth of these characteristics have been adequately verified, with the exception of a statement qualifying item 2. Baade and Zwicky stated that the observed amount of energy radiated by a supernova should involve "a considerable fraction of the mass of the star". Mrs. Gaposchkin [10] finds that the loss due to radiation, which is of the order of 10^{27} grams during an outburst, is but an insignificant fraction of the total mass; and Landau [11] finds, on the neutron-core hypothesis, that the radiated energy would involve only about two per cent of the mass.

The spectra, although difficult of interpretation, do not appear to differ greatly from the spectra of normal novæ in regard to the elements represented. The widths of the bands, however, imply ejected-material velocities from three to five times as great as for ordinary novæ.

The observable features of the spectra seem to indicate that the surface temperatures cannot exceed 50,000° and are probably smaller. If the temperatures were as high as those predicted, the lighter atoms would be stripped of electrons and the spectra would consist mainly of soft X-rays which would not be observable because of the interference of our own atmosphere.[12] An analysis by Minkowski of all the supernovæ spectra obtained at Mount Wilson is promised shortly [13] and should

[9] *P. A. S. P.*, 50, 215, 1938.
[10] *P. N. A. S.*, 22, 334, 1936.
[11] *Nature*, 141, 334, 1938.
[12] *P. N. A. S.*, 22, 335, 1936.
[13] *Mount Wilson Contribution 602*, in press.

greatly increase the present knowledge of the physical conditions in these objects.

There is no convincing direct evidence that supernovæ are a source for cosmic rays. McCrea[14] finds, on "reasonable assumptions," that supernovæ could emit cosmic rays, though the greater part of the energy loss would occur otherwise than in light radiation. On the other hand, Vallarta, an authority on cosmic rays, considers this detail of Zwicky's theory questionable. Observations of cosmic rays at the times of novæ outbursts have as yet yielded negative results concerning variations in intensity or in directional effects.

The neutron-core theory for supernovæ forms a part of Dr. Whipple's discussion and will not be treated here. The remainder of the present paper will be devoted to a presentation of the observational data available for the determination of the absolute magnitudes of supernovæ and the amounts of energy involved in their radiation. The absolute magnitudes were determined from the apparent magnitudes and the distance moduli of the parent-nebulæ. An estimate of the amount of energy radiated in a given time was obtained from the light curves.

The Light Curves. Baade[15] has given a complete account of, and has revised the magnitudes for, the eighteen previously published supernovæ that were known to him. Harvard plates have been searched for additional images of all the twenty-one extra-galactic supernovæ. In very few instances has a supernova been observed definitely at maximum brightness. Long series of post-maximum observations are also rare. Combining published data with the new estimates we have been able to construct, for eight of the supernovæ, fairly detailed light curves covering a period of over one hundred days since the observed maximum. The curves are shown in Fig. 2; dots indicate material obtained from published sources, and circles represent the recent Harvard estimates. Data for the less completely observed supernovæ are shown in Fig. 3. All the curves of Fig. 2, except that for the problematic SS Ursæ Majoris which remained at observed maximum over seventy days, indicate sharp maxima. The ascending branch of the light curve is in no instance well defined. Observations of the three supernovæ in NGC 5253 (Z Centauri), 4273, and 4303 show that the rise to maximum occurred within an interval of twenty days preceding the observed maximum, but the data are too scanty to define the time of rise more exactly. Somewhat better data are available for S Andromedæ, which apparently attained its full maximum brightness within the one day from August 16

[14] *Nature*, **135**, 371 and 821, 1935.
[15] *Cf.* Ref. 8.

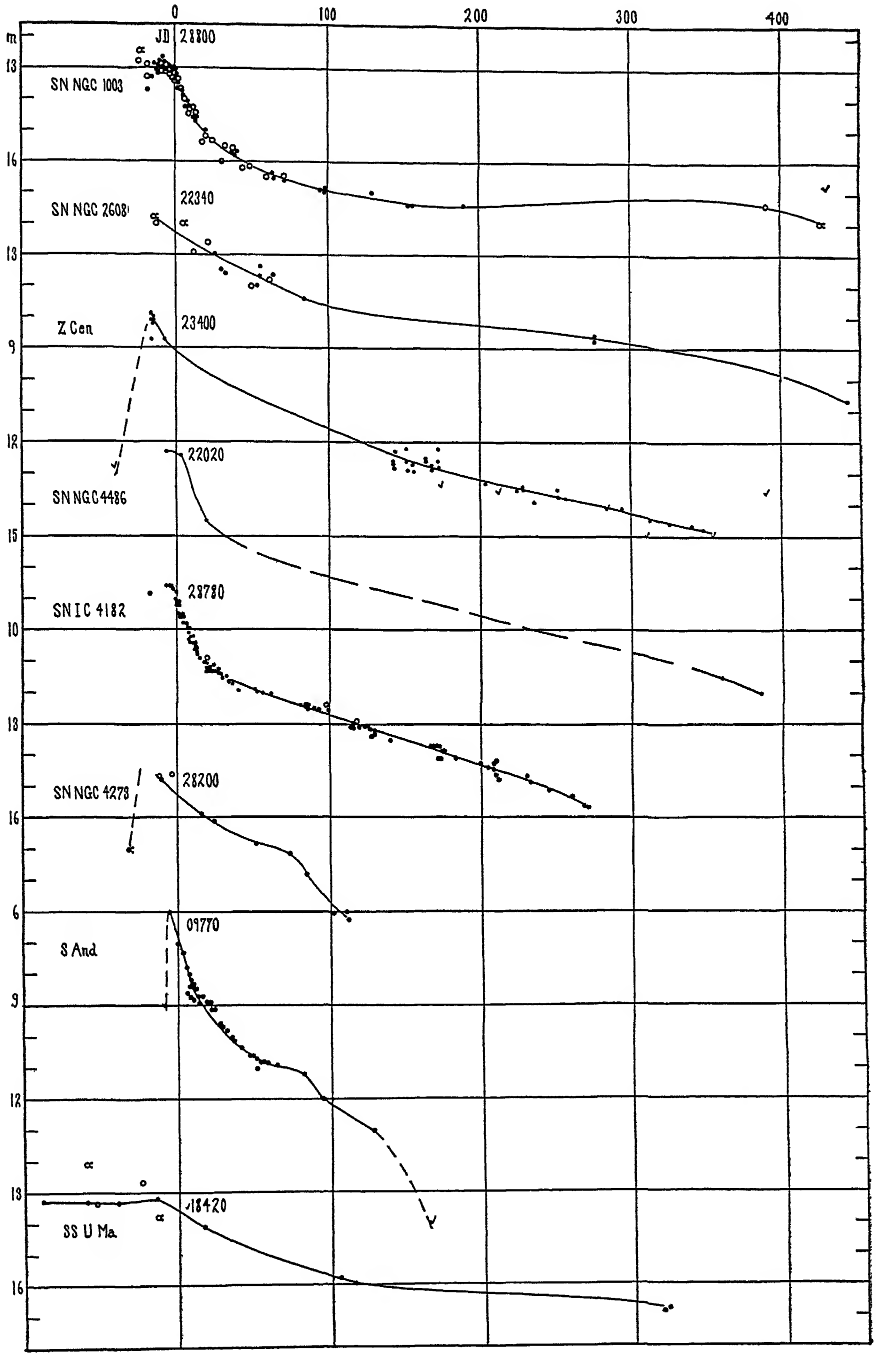


Fig. 2. Light curves of the best observed supernovæ.

to August 17, 1885. Three other novæ to show fairly rapid increases in brightness were SN IC 4719 which varied from $15^m.1$ to $13^m.7$ in four days, SN NGC 2535 which increased from $16^m.4$ to $14^m.7$ in eight days, and SN NGC 4636 which brightened from 14^m to $12^m.5$ in one day. The rate of decrease in brightness after maximum is rapid at first, becoming more gradual within a few weeks.

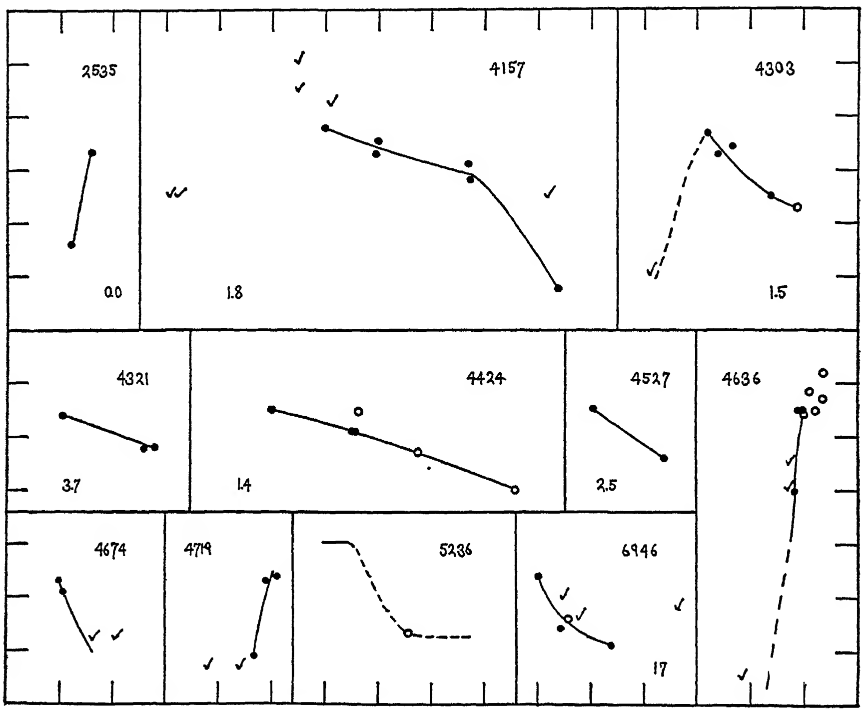


FIG. 3. Partially observed light curves. Coordinate scales indicate one magnitude and 20-day intervals. Numbers in upper parts of the diagram refer to the NGC or IC objects; those in lower corners indicate number of magnitudes by which Baade extrapolated to maximum. The individual observations for SN NGC 5236 were not available: Baade (ref. 8) has indicated the trend of the curve from unpublished material by Lampland.

The supernova light curves appear to resemble those of typical ordinary novæ (Fig. 4) in the probable steepness of rise to maximum and in the evidence for more rapid fading within the first twenty to forty days than in their later history. During approximately the first thirty days the rate of decline of supernovæ, $0^m.1$ per day, is about twice the average for the ordinary novæ found in the Andromeda Nebula, $0^m.05$ per day.[16] The supernova curves, however, give little indication of the

[16] Hubble, *Ap. J.*, 69, 139, 1926.

rapid post-maximum fluctuations found in many ordinary novæ (*e.g.*, Nova Pictoris 1925).

Although exhibiting common features, the curves for the supernovæ show marked differences in slope after the first period of rapid decline; again with the exception of that for SS Ursæ Majoris, they can all be arranged in a continuous sequence determined by the magnitude gradients after the first hundred days following maximum. The supernova in NGC 1003 shows the flattest curve, S Andromedæ the steepest. The

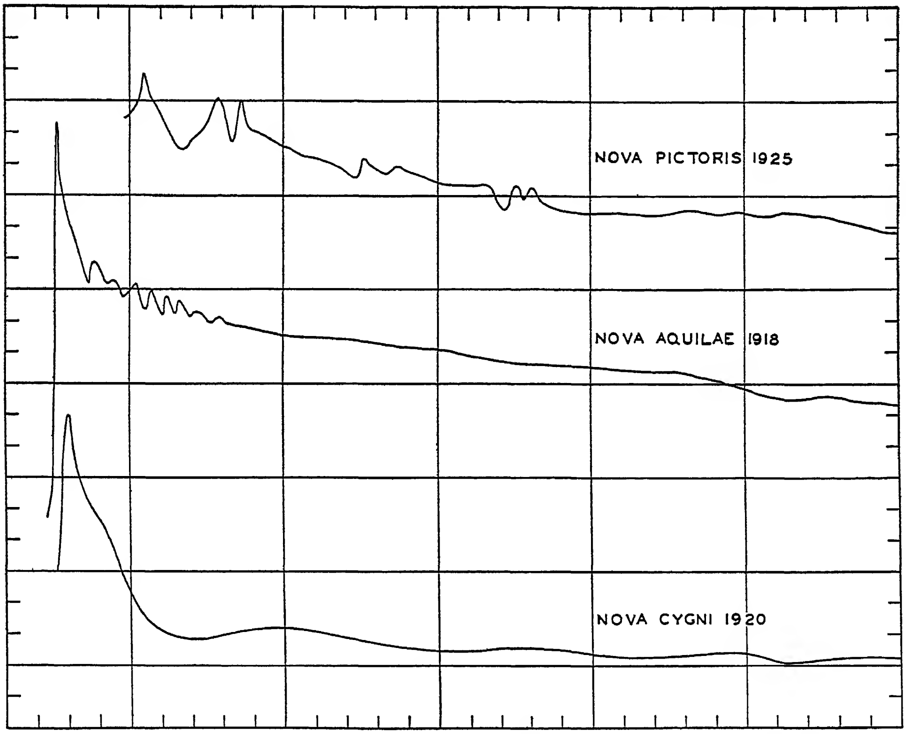


FIG. 4. Light curves of typical ordinary novæ. Coordinate scales indicate 20-day and one-magnitude intervals (for data, see Campbell, H. B. 890, 1932).

differences are apparent in Table I which gives the magnitude decrease for each supernova within one-, two-, three-, and four-hundred days after the highest observed maximum. SN NGC 1003 varied only 4.5 magnitudes, whereas S Andromedæ varied over ten magnitudes in two hundred days.

Baade has fitted the mean of the light curves of SN NGC 1003 and SN IC 4182 to the estimates of other supernovæ, in order to extrapolate magnitudes at maximum for those insufficiently observed. His pro-

cedure yields a good statistical mean absolute magnitude. For an individual supernova the extrapolated magnitude may, however, have but limited significance when only two or three observations are available (*cf.* Fig. 3). We note, for example, that the curves for S Andromedæ and SN NGC 4273 (Fig. 2) show humps between eighty and one hundred days following maximum. The gradients just after the humps are comparable with the gradients just after maximum. If the two or three observations, from which magnitudes at maximum were extrapolated, happened to correspond to a region of the curve just past such a hump, the extrapolated magnitudes might be as much as four magnitudes in error.

An additional useful criterion in the extrapolation of a magnitude at maximum might be the integrated magnitude of the nebula in which a supernova appeared. Baade has found that the average difference between the magnitude at maximum and the integrated magnitude of the nebula is $+0^{M}.34$, but the spread is eight magnitudes (see Table II, column nine).

TABLE I

SN	Magnitude Decrease of Supernovæ			
	100^d	200^d	300^d	400^d
	m	*m*	*m*	*m*
1003	4.0	4.5	4.5	4.5
2608	2.5	3	4	5
Z Cen	3.5	5	6	7
4486	4 ?	5 ?	6 ?	8 ?
4182	4	6	7	—
4273	4	>4.5	—	—
S And	6	>10	—	—
SS UMa	2.5	3	3.5	—

Table II gives an abstract of the data and results published by Baade. The observed magnitudes at maximum, given in the fourth column, are only for the curves that he extrapolated by less than half a magnitude; the magnitudes in the fifth column were extrapolated by the amounts given in the column following. The mean value of the absolute magnitude, obtained from the apparent magnitudes and the distance moduli, is $-14^{M}.3$. The mean for only the five fairly reliable (unextrapolated) observed values is likewise $-14^{M}.3$.

The Integrated Light Curves. Supernovæ do not radiate like black bodies, nor do their spectra, consisting of wide emission bands, resemble the solar spectrum. In order to obtain a *lower limit* to the amount of

TABLE II

NGC or IC	Type Neb.	Dist. Mod.	Obs. m_{max}	Extrap. m_{max}	Obs - Extrap.	M_{max}	m_{neb}	$m_{nova}-m_{neb}$	M_{neb}
224......	Sb	22.2	7.2	—	0	−15.0	4.5	+2.7	−17.7
1003......	Sc	26.8	12.8	—	0	−14.0	13.1	−0.3	−13.7
2535......	SBc	—	—	14.7	0	—	13.7	+1.0	—
2608......	SBc	—	—	11.0:	1.9	—	13.6	−2.6	—
2841......	Sb	—	—	—	—	—	10.6	—	—
4157......	Sc	—	—	14.4	1.8	—	12.0	+2.4	—
4182......	Sc	24.8	8.6	8.2	0.4	−16.6	13.5	−5.3	−11.3
4273......	Sc	26.7	14.8	14.4	0.4	−12.3	12.4	+2.0	−14.3
4303......	SBc	26.7	—	12.8	1.5	−13.9	10.4	+2.4	−16.3
4321......	Sc	26.7	—	11.9	3.7	−14.8	10.5	+1.4	−16.2
4424......	SBb	26.7	—	11.1	1.4	−15.6	12.5	−1.4	−14.2
4486......	Eo	26.7	12.3	12.0	0.3	−14.7	10.1	+1.9	−16.6
4527......	Sc	26.7	—	13.0	2.5	−13.7	11.3	+1.7	−15.4
5236......	Sc	24.8	—	—	—	—	8.8	—	—
5253......	Irr	—	8.0	—	0	—	11.0	−3.0	—
5457......	Sc	23.8	—	—	—	—	8.9	—	—
6946......	Sc	25.3	—	12.9	1.7	−12.4	11.1	+1.8	−14.2
Mean of All............						−14.3		+0.34	−15.0
Mean of Reliable (unextrapolated)................						−14.3		−0.15	

radiant energy emitted by a supernova, we have however assumed a spectral energy distribution similar to that of the sun. For six of the supernovæ with well-determined light curves we have distance moduli and can therefore convert the apparent into absolute magnitudes. Assuming

$$I/I_{\odot} = 0.4\ (M_{\odot} - M_{\text{nova}})$$

(where I is the intensity of the radiation of the supernova; $I_{\odot}$, the intensity of the solar radiation = 1.2×10^{41} ergs/year; and $M_{\odot} = +5.3$), we have found the values ΣI given in Table III for the total amount of energy radiated in the photographic region, within one hundred days and within one year of maximum. Since the tabulated

TABLE III

SN	M_{max}	$I_{max}/I_{\odot}$	$\sum\limits_{100^d} I$	$\sum\limits_{\text{year}} I$
NGC 1003	−14.0	5.3×10^{7}	5.0×10^{47}	5.8×10^{47} ergs
NGC 4486.... .	−14.4	7.6×10^{7}	5.8×10^{47}	6.3×10^{47}
IC 4182....... ...	−16.2	4.0×10^{8}	3.6×10^{48}	
NGC 4273.......	−12.4	1.2×10^{7}	1.2×10^{47}	
S And......	−15.0	1.3×10^{8}	3.6×10^{47}	
SS UMa..........	−10.6	2.3×10^{6}	6.7×10^{46}	8.8×10^{46}

values are lower limits, the value of 10^{48} to 10^{49} ergs, predicted by Zwicky, is probably a good approximation.

Both theoretically and observationally the study of supernovæ is still in its early beginnings. Astrophysicists may at times find themselves in disagreement with various phases of Baade's and Zwicky's theory as to the cause of supernova outbursts. Zwicky's list of seven supernova characteristics serves, however, as an inciting plan of research both for the observer and the theorist. Both authors are to be heartily congratulated on their successful efforts in accumulating data on this spectacular type of star, in quantities that have not been obtained before. With the continuation of such efforts we may expect eventually to know a great deal more than at present about the true nature of supernovæ.

THE MATERIAL OF INTERSTELLAR SPACE

THEODORE DUNHAM, JR.

Mount Wilson Observatory, Carnegie Institution of Washington

(*Read February 17, 1939, in Symposium on Progress in Astrophysics*)

UNTIL a very few years ago it was generally believed that almost all of the material in the universe was in the stars, and that clouds of dust, wandering meteors and occasional planets made up a very unimportant contribution to the total mass. Evidence has been accumulating, however, at first gradually, and now more rapidly, to indicate that the space between the stars contains not only cosmic dust and meteors, but also free atoms and electrons in such great amounts that the combined mass of this almost invisible material may nearly equal or even exceed that of all the stars.

An upper limit for the total density of material of all kinds in space has been set by Kapteyn and Oort on the basis of the observed velocities of the stars. Their value for this limiting density is close to 6×10^{-24} grams per cm^3. The visible stars account for very nearly half of this value, leaving an upper limit of approximately 3×10^{-24} grams per cm^3 for the total density of meteors, dust and gas.

Meteors are made known to us when they fall from outer space into our atmosphere and flash for an instant before they are consumed. Their mass and numbers have been estimated by various workers with reasonable accuracy. More than a billion meteors strike the earth each day, and this figure indicates that on the average they are between 50 and 100 kilometers apart in space. The mass of an average meteor is a little more than a tenth of a milligram, which leads to a mean density in space of about 5×10^{-25} grams per cm^3.

Cosmic dust reveals its existence in three different ways. Occasionally, as in the Pleiades, bright stars illuminate local clouds so intensely that they may be seen directly on photographic plates. More often these clouds are detected when they block out the light of stars behind them, as in the great coal sack in the southern Milky Way. And finally, dust in finely divided form reddens the light of distant stars which has passed through it, just as light from the setting sun is reddened by a long journey through our own atmosphere.

Much may be learned about the size of the particles of cosmic dust from a study of the law connecting the absorption coefficient (k) with the wave-length of light. The observations agree reasonably well and indicate that the absorption coefficient varies very nearly inversely with the wave-length ($k \sim \lambda^{-1}$). If the absorption were caused by particles much smaller than a wave-length of light, by atoms or molecules for example, the coefficient would vary inversely with the fourth power of the wave-length ($k \sim \lambda^{-4}$), while if it were caused by particles much larger than a wave of light the coefficient would be constant throughout the spectrum (k = constant). Clearly the particles must have average dimensions comparable with the length of a wave of light. Detailed study of the absorption law indicates that the mean diameter of cosmic dust particles is a little less than 10^{-4} mm, and that the mean density of this dust is approximately 10^{-26} grams per cm^3. Neither meteors nor dust appear to contribute nearly as much mass to our stellar system as do the stars.

When a cloud of gas consisting of free atoms is close to a star it may be rendered visible by fluorescence as in the case of the Orion Nebula and other gaseous nebulæ. Otherwise the detection of atoms or molecules in space depends entirely on the spectroscope which shows absorption lines in the spectra of distant stars whenever the light has passed through an atomic or molecular cloud on its way from those stars to the earth.

The appearance of an interstellar line is influenced by at least five factors: (1) The atom from which it arises must be reasonably abundant in space; (2) only the most important lines absorbed from the lowest atomic energy level in each atom are likely to be observed; (3) these can be observed only if they lie between λ3000 and λ10,000; (4) the lines of an element have a much better chance of being detected if there are only one or two strong ultimate lines than if the effort is divided between a larger number of lines from the ground state; (5) a high ionization potential for the ion concerned, or for the next more highly ionized ion is a circumstance favoring the visibility of absorption lines, because unless ionization is blocked most of the atoms will exist in the form of higher ions and the effective abundance of the ion under consideration will be greatly reduced. On the basis of these requirements, the following atomic absorption lines might be expected to show in the spectra of distant stars:

**Na* I	3302, 3303, 5890, 5896
Al I	3944

* Detected up to February 1939.

**K*	I	7664, 7698
**Ca*	I	4226
**Ca*	II	3933, 3968
Sc	II	3642, etc.
**Ti*	II	3073, 3229, 3242, 3383
Fe	I	3719, etc.
Sr	II	4077, 4215
Ba	II	4554, 4934

Interstellar lines are most easily detected in stars of early type (O to B3), having few absorption lines of their own. They are usually much sharper than the stellar lines and their position ordinarily indicates a different radial velocity from that given by the stellar lines.

Hartmann [1] established the existence of interstellar *Ca* II in 1904 in the spectrum of δ Orionis. The H and K lines stood out among all other lines in the spectrum of this spectroscopic binary because they took no part in the periodic radial velocity changes. In 1919 Miss Heger [2] at the Lick Observatory discovered the D lines of *Na* I, and not long afterward the ultra-violet pair of sodium lines was photographed by Wright.

The use of an aluminized plane grating, ruled by R. W. Wood, with off-axis Schmidt cameras [3] at the coudé focus of the 100-inch telescope has led to the discovery of interstellar *Ti* II,[4, 5] *Ca* I,[6] and *K* I.[6]

The failure to detect *Al* I, *Sc* II, *Fe* I, *Sr* II and *Ba* II up to the present time is not remarkable. Aluminum is mostly in the ionized condition. Scandium, strontium and barium are relatively rare. In the case of scandium and iron, moreover, there are several important transitions from the ground state, so that none of them is as strong as it would be if the effort were not divided.

In the case of *Ti* II, the lowest term (^{4}F) consists of four sub-levels, and it is of considerable interest to notice that, although the lines arising from the lowest sub-level are easily seen, no trace can be found of interstellar lines from the next higher sub-level. This level has an excitation potential of only 0.012 volt and in laboratory sources the lines from this state are actually stronger than those from the lowest level. A consideration of the structure of the *Ti* II spectrum,[5] shows that this situation can exist only if there are spontaneous downward transitions between sub-levels in the ^{4}F term. This astrophysical observation appears to be the only direct evidence for the existence of transitions of this type.

Several unidentified lines of interstellar origin have recently been found at Mount Wilson. Four of these, in the yellow and red,[7] are

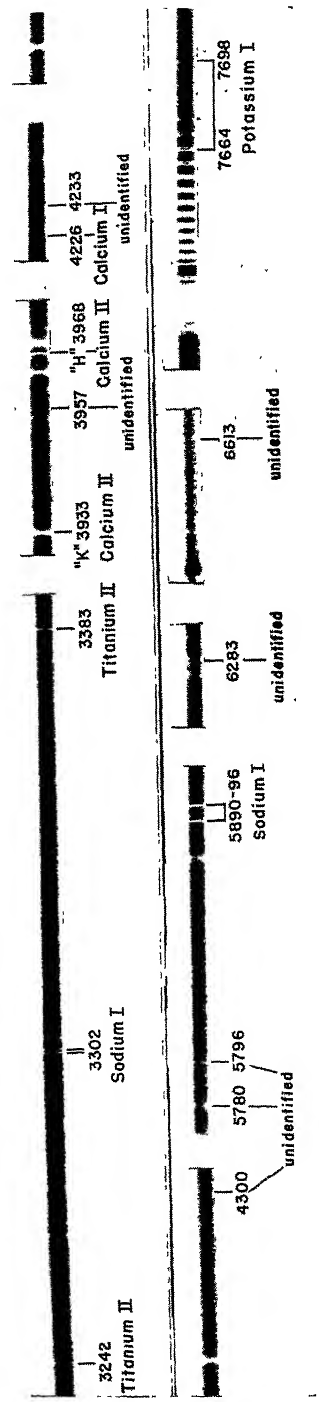

3242
Titanium II
3302
Sodium I
3383
Titanium II
"K" 3933
Calcium II
3957
unidentified
"H" 3968
Calcium II
4226
Calcium I
4233
unidentified
4300
5780
5796
unidentified
5890-96
Sodium I
6283
unidentified
6613
unidentified
7664
7698
Potassium I

FIG. 1

broader than other interstellar lines. Swings [8] has considered the possibility that they represent a blending of the first few rotational lines in a molecular band, while Merrill [9] has advanced the hypothesis that they may be due to absorption by solid particles in space.

Three unidentified interstellar lines in the violet region of the spectrum,[6] at λ3957, 4233 and 4300 are extremely sharp. Their intensities do not appear to be correlated either with one another or with the intensities of *Na* I or *Ca* I. In ζ Ophiuchi the strength of λ4233 is surprisingly high, being almost half as strong as the H line of *Ca* II, and much stronger than λ4300. In other stars the ratio 4233/4300 is reversed. These lines cannot be identified with the ultimate lines of any of the elements. It is possible that they represent absorption from the lowest rotational levels of molecular bands, and it also seems possible that they may be upward transitions from atomic metastable states from which the downward transitions are strongly forbidden. Such lines would have been observed in the laboratory before this, unless they belong to an atom ionized more than once.

The entire interstellar spectrum, as it is known at the present date, is shown in Fig. 1. Several stars are represented, the best one being selected to illustrate the lines in each region.

One of the most interesting of all the interstellar lines is that of neutral calcium at λ4226, because it provides one instance of an element having two stages of ionization available for study in the interstellar spectrum. The observed ratio of *Ca* I to *Ca* II makes it possible to calculate the degree of ionization in space and, if the quality and intensity of interstellar radiation can be estimated, we may compute the density of electrons in space and the numbers of other observed elements in each stage of ionization.

A word of caution is needed at this point. The entire argument in the remaining portion of this paper depends on the intensity of *Ca* I 4226. Unfortunately the line is a singlet and there are no other transitions from the ground state strong enough to serve as checks on the identification. The agreement in wave-length with the laboratory value is excellent, but the line is exceedingly faint and it must be remembered that there are two lines of unidentified origin and of similar interstellar appearance within 100 Angstroms of λ4226. There is a definite possibility that the line in question is not absorbed by neutral calcium, but that we are dealing with a chance coincidence with another unidentified line. This seems unlikely but must be kept in mind. Another possibility is that the line arises from *Ca* I in an extended atmosphere. The velocity derived from λ4226 agrees very well with that

derived from the interstellar lines of *Ca* II and *Na* I in the spectra of two stars and this velocity is definitely different from that given by the stellar lines. This evidence suggests that the line is truly interstellar in its origin, but observations on a larger number of stars will be required before the matter can be considered as finally settled.

The observations, on which the calculations which follow are based, were made with the 32-inch and 73-inch focus Schmidt cameras and the second-order grating ruled by R. W. Wood. The dispersion is approximately 10.2 and 4.5 A/mm respectively for the two cameras. Emulsions of high contrast were required even to show the existence of some of the fainter interstellar lines, especially the one at $\lambda 4226$. This line shows beyond any doubt in the spectrum of χ^2 Orionis, and it is believed that its intensity has been measured with sufficient dependability to give at least the right order of magnitude in calculations based upon it. It is a fortunate circumstance that faint lines fall on the portion of the curve of growth where the slope is 45° so that an error in measurement is much less serious than it is in the case of somewhat stronger lines which fall on the portion of the curve where the slope is much less. In the spectrum of 55 Cygni the line at $\lambda 4226$ can barely be seen and, although the wave-length measures are accordant, measurements of intensity are highly unreliable. The line has not yet been detected with complete certainty in any other star. In what follows the discussion will therefore be limited to the case of χ^2 Orionis, which is definitely an exception in the matter of its interstellar lines.

The intensities given in Table I are preliminary values and will doubtless be improved when better photographs are available. The measures of the potassium line are particularly uncertain, partly because the dispersion in the first-order infra-red with the 32-inch camera is only about 20 A/mm, and partly because it was necessary to derive the photometric calibration indirectly by using lines in the A-band of oxygen.

The interstellar K-line of *Ca* II in χ^2 Orionis is affected by a blend in the wing on the long wave-length side. There is some indication that this has the characteristics of an interstellar component, but an examination of the region near H gives no clear confirmation that this is the case. Measures of K based on lower dispersion may be too high by about 10 per cent on account of this blend. The value for this line in Table I is an average of the intensity given by Merrill, Sanford, Wilson and Burwell [10] and the intensity derived from coudé plates after allowing for the blend.

Intensities of Fraunhofer absorption lines are expressed in terms of a

TABLE I

INTENSITIES OF INTERSTELLAR LINES IN χ^2 ORIONIS

Atom		λ	f	F	log F	log $Nf\lambda$	N
Na I	D_1	5896	0.33	66	1.82	9.34	5.0×10^{-8}
Na I	D_2	5890	0.66	78	1.89	9.64	5.0×10^{-8}
Na I	D_1'	3303	0.0047	12	1.08	7.24	5.0×10^{-8}
Na I	D_2'	3302	0.0094	18	1.26	7.54	5.0×10^{-8}
Ca I		4226	1.00	3.3	0.52	6.57	4.0×10^{-11}
Ca II	K	3933	0.66	95	1.98	9.91	1.40×10^{-7}
K I		7698	0.33	21	1.32	7.67	8.3×10^{-10}
Ti II		3383	0.43	21	1.32	7.67	1.4×10^{-9}
Ti II		3242	0.27	15	1.18	7.37	1.2×10^{-9}

unit (referred to as an F-unit) which is equal to the energy contained in a section of continuous spectrum the length of which is one millionth of the wave-length. This unit (formerly referred to as a micro-wave-length) was proposed by the present author [11] in 1933, and has marked advantages when lines which differ greatly in wave-length must be compared. It is easy to show that no shift in the curve of growth is required, when different regions of the spectrum are involved, if intensities are expressed along the axis of ordinates in units of $10^{-6}\,\lambda(=\mathrm{F})$ and if $Nf\lambda$ is plotted along the axis of ordinates. A unit of intensity proportional to wave-length has been used by Allen [12] and by Williams,[13] and its theoretical advantage has been discussed by Menzel.[14] In order to convert F-units into equivalent Angstrom units it is only necessary to multiply by the wave-length and divide by one million.

$$\text{E.A.} = \text{F} \times 10^{-6} \times \lambda$$
$$(1\text{F} = 10^{-6} \times \lambda)$$

In Table I, the first column gives the atomic symbol and state of ionization of the spectral line concerned, followed by its name, if any, in the solar spectrum. The second column gives the wave-length and the third the value of f, the oscillator strength. The fourth column gives the measured intensity of the line in the spectrum of χ^2 Orionis in F-units. The logarithm of the intensity is entered in the fifth column. The sixth column gives the abscissa (log $Nf\lambda$) for the point corresponding to each spectral line when plotted on the curve of growth. This will be explained later. The last column gives the number (N) of atoms of the species concerned in an average cubic centimeter of the

path between the earth and χ^2 Orionis. The derivation of this value will also be explained later.

It would be of great interest to know the total number of atoms of *Na, Ca, K* and *Ti* per cubic centimeter in interstellar space. But each of these atoms exists, to an appreciable degree, in two or three stages of ionization, all in equilibrium with one another. In the case of calcium, for example, the following equilibria are maintained:

$$Ca\ \text{I} \rightleftarrows Ca\ \text{II} \rightleftarrows Ca\ \text{III} \rightleftarrows Ca\ \text{IV}.$$

Except in the case of calcium (where both *Ca* I and *Ca* II are observable) only one stage of ionization is visible (*Na* I, *K* I, and *Ti* II). If we are to count all the atoms of each of these four elements we must count the invisible ions, as well as those which can be directly observed. To do this we must first know the ratios between the numbers of successive ions, then compute the actual number of each type of ion by using the one observed ion as a reference, and finally add all types of ions to get the complete atomic count.

There is in space a definite equilibrium attained between the rate at which ultra-violet light from all the stars combined is tearing electrons off from normal atoms and the rate at which the resulting ions recapture free electrons. We have

$$\begin{array}{ccc}
\text{Photons} + & Ca\ \text{I} \rightleftarrows Ca\ \text{II} & + \text{Electron} \\
| & | \qquad\quad | & | \\
I_\lambda & \underbrace{[Ca\ \text{I}] : [Ca\ \text{II}]} & N_e \\
(a) & (b) & (c)
\end{array}$$

In this photochemical reaction we may attempt to estimate (*a*), the intensity of ionizing radiation, from our knowledge of the temperatures and numbers of the stars of each degree of brightness. In the case of calcium (*b*), the ratio $\frac{Ca\ \text{I}}{Ca\ \text{II}}$, is directly observed. When the ionization potential is known these data permit an immediate calculation of (*c*), the number of electrons per cubic centimeter. This determination of the electron density is an interesting result in itself, but its particular importance is that it allows all other ionization ratios to be computed, by working the equations backward. We start with (*a*), the density of photons, and (*c*), the density of electrons, and then by using the appropriate ionization potential we solve for (*b*), the density ratio of two consecutive ions.

It was Eddington [15] who first considered the question of interstellar ionization. He concluded that, if all the electrons in space come from

sodium and calcium atoms, the level of ionization must be extremely high. He obtained a ratio *Ca* II : *Ca* I = 300,000 : 1. If the ratio were as high as this there would be little hope of detecting the absorption line of *Ca* I at λ4226. At about the same time Gerasimovič and Struve [16] studied the question and came to similar conclusions as to the high level of ionization in space.

Computations based on Mount Wilson measures of the calcium ionization ratio in the path between the earth and χ^2 Orionis were made a little more than a year ago.[17] These calculations depended on a simplified picture of the interstellar radiation in which no allowance was made for absorption in the ultra-violet. An upper limiting value of 20 electrons per cubic centimeter was derived, which was many times as great as the number of all known positive ions combined. The only possible source for so great a concentration of electrons is the ionization of a somewhat greater number of free hydrogen atoms. Heavier atoms would raise the average density of space even more than does hydrogen above the Kapteyn-Oort limit.

Langer [18] has computed the level of ionization, starting with Struve and Elvey's [19] result that there is about one hydrogen atom per cm^3 in certain galactic nebulæ. He assumes that the hydrogen is mostly ionized and that there must therefore be approximately one electron per cm^3. Langer's results for atomic concentrations differ considerably from ours, largely because a different assumption was made as to the quality of the interstellar radiation.

In an extremely interesting paper Struve [20] has recently discussed the situation on the basis of our detection of the *Ca* I line at λ4226. He assumed an intensity of 0.01 E.A. for this line and derived a ratio for *Ca* II : *Ca* I = 150 : 1, using the curve of growth which Wilson and Merrill [21] obtained from the D lines of sodium by combining measures from a large number of stars. This ratio is more than twenty times smaller than our value derived from a curve of growth based on the four interstellar sodium lines in the spectrum of χ^2 Orionis. The statistical weights were not included in the ionization equation used in the calculation and it so happens that this fact reduces the result for the number of electrons per cm^3 to 30, a value about three times as large as that derived in the present work. Struve emphasizes the great amount of hydrogen implied by so high a concentration of electrons.

Before measurements of absorption line intensities can be applied to a study of ionization equilibria a relation must be established between the number of absorbing atoms and the intensity of the resulting absorption line. This relation, known as the "curve of growth," is not

the same under different physical conditions. For weak lines, when the number of active atoms is small, the absorption is directly proportional to the number of atoms, so that when log F (absorption) is plotted against log **N**f (the effective number of atoms) the curve is a straight line with a 45° slope. But at a certain point, further along the curve if thermal or turbulent motion of the atoms is increased, saturation begins to set in and the curve flattens until its slope is greatly reduced. Finally, as **N** increases, the wings of the lines beyond the Doppler region begin to be effective, and from that point on the absorption increases in proportion to $\sqrt{\mathbf{N}}$.

In interstellar space conditions are so complex and so little known that it would be highly dangerous to use a curve of growth based on any assumed set of physical conditions (temperature and turbulence). Wilson and Merrill's [21] curve represents average conditions in space, but may not apply to the path between the earth and one particular star. The safest course is to establish an empirical curve from interstellar lines in the spectrum of the star concerned. In the case of some of the more distant stars, such as χ^2 Orionis, this is fortunately possible, because the relatively weak ultra-violet sodium lines (referred to as the D′ lines) at λ3302 and λ3303 may be combined with the much stronger D lines of sodium in the yellow, at λ5890 and λ5896, to give a relation between effective numbers of atoms and observed units of absorption. The f-values for the four sodium lines were taken from the measures of Filippov and Prokofjew [22] and are given in the third column of Table I.

The observed intensities of the four interstellar sodium lines (log F) are plotted against log **N**$f\lambda$ in Fig. 2. At this point the logarithmic scale of abscissæ is entirely arbitrary, since it is only the *differences* in the values of log **N**$f\lambda$ which are known—f is known for each line, but **N** is completely unknown, although it is obviously the same for all four lines.

The slope of the curve in the region corresponding to moderately weak lines is given by a line through the two D′ lines. This is not 45°, but is much greater than the slope through the D lines. The slope must become 45° for lines a little fainter than the D′ lines, as experience with the D-line ratio in nearby stars indicates. It is not difficult to see approximately how the transition is accomplished. The 45° line forming the lowest portion of the curve of growth in Fig. 2 represents what is believed to be the most probable position of this part of the curve. It may actually lie a little to one side or the other of this position.

The absolute value of the logarithmic scale of abscissæ is now de-

termined. A relation of fundamental importance and reliability exists [23] between the total absorption of very faint lines and the number of atoms concerned in their production:

$$F = \frac{\pi e^2}{mc^2} \cdot 10^6 \cdot Nf\lambda$$

$$= 8.8 \cdot 10^{-7} \cdot Nf\lambda.$$

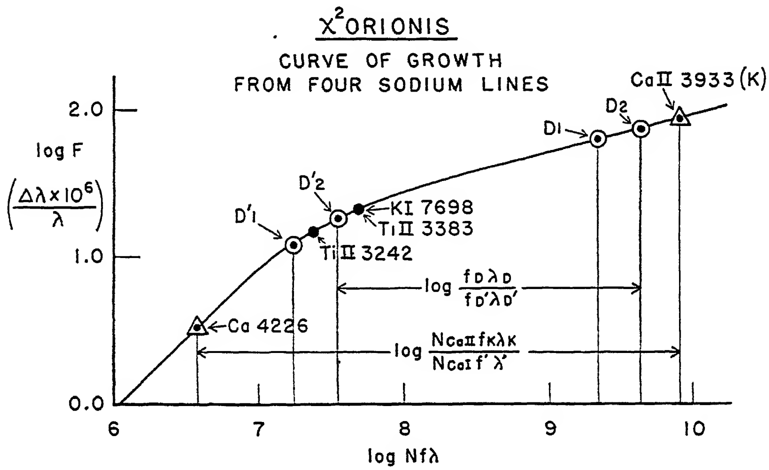


FIG. 2

From this relation it follows that when log F = 0.00 (i.e. F = 1.00), log $Nf\lambda$ = 6.055. The scale of ordinates has been adjusted in Fig. 2 to make the straight portion of the curve cross the axis of abscissæ at this point.

On this curve, built up to fit the four sodium lines, have been plotted points whose ordinates represent the observed absorption intensities (log F) of the lines of other elements (*Ca* I, *Ca* II, *K* I and *Ti* II). The abscissa corresponding to each point gives directly a value of $Nf\lambda$ for the line concerned. Since f and λ are known (see Table I) this gives immediately a value of **N**. This is the total number of atoms (in the stage of ionization considered) between the earth and the star.

The distance of χ^2 Orionis is not known with great accuracy. It is a super-giant star of class B1. If we assume an absolute magnitude of − 5.0 and make a reasonable assumption for the influence of general absorption in space, the distance is found to be about 725 parsecs = 2.23×10^{21} centimeters. Dividing the number of atoms (**N**) of each type by this value for the length of path, gives the number (N) of

atoms of that type in an average cubic centimeter along the path. This result is entered in the last column of Table I.

The values for the concentration of neutral and of singly ionized calcium give the first ionization ratio of calcium:

$$\frac{Ca\,\mathrm{II}}{Ca\,\mathrm{I}} = \frac{1.40 \times 10^{-7}}{4.0 \times 10^{-11}} = 3500.$$

If the quality and quantity of interstellar radiation to which the *Ca* I atoms are subjected can be estimated, it will be possible to compute what concentration of electrons must exist so as to drive back the ionization at a rate which will maintain a balance in the reaction at the point where the ratio of *Ca* II to *Ca* I has its observed value of 3500.

The integrated radiation of all the stars has been estimated as accurately as the data appear to warrant. First, on the assumption that there is no absorption in space, the stars of various spectral types were grouped into ten different temperature classes. The mean temperatures given in Table III were assumed for each class and the number of equivalent stars of 0.00 magnitude in this class was determined (column 3). The dilution factor (δ) for black body radiation corresponding to this number of 0.00 magnitude stars was computed, and from it the actual density of radiation at 16 wave-lengths between λ333 and λ10,000. At each wave-length the contributions from the ten temperature classes was summed to get the total radiation at that wave-length.

In order to estimate the influence of selective absorption on the degree of interstellar ionization, the same computation was carried out with the assumption that there is a general photographic absorption, $A_{pg} = 1.20$ mag/kps, and that there is a color excess, $E_{55/44} = 0.40$ mag/kps, equal in all directions. The mean distance of stars in each temperature class for each magnitude was estimated from an assumed absolute magnitude, and then the appropriate amount of absorption was applied at each wave-length.

The density of radiation per cm³ is given in Table II in terms of a unit = 10^{-20} erg/cm³/A. This table is based on a large amount of numerical work, and while an effort has been made to avoid errors, the results are of a preliminary character and it is more than likely that some of the values will be changed appreciably when the computations have been completely verified. It is felt that this summary may be worth including at this time merely to illustrate the extent to which one O-star dominates the ultra-violet. By contrast, Sirius contributes a very moderate fraction of the radiation, even near its energy maximum.

TABLE II

DENSITY OF INTERSTELLAR RADIATION (ERGS/CM^3/A. $\times 10^{20}$)

λ	No Absorption			With Absorption		
	All Stars	ζ Puppis	Sirius	All Stars	ζ Puppis	Sirius
333	1,580	1,000	1.61×10^{-10}	65.5	60.2	1.61×10^{-10}
400	2,100	1,215	0.010×10^{-5}	142.0	124.3	0.0103×10^{-5}
500	2,420	1,205	4.35×10^{-5}	265	203.5	4.35×10^{-5}
667	2,610	884	0.013	464	245.0	0.013
833	2,960	584	0.317	728	221.4	0.317
1,000	3,400	379	2.24	1,080	214.8	2.24
1,250	4,060	215	12.9	1,550	116.2	12.9
1,667	4,590	86	53.3	2,250	59.5	53,3
2,000	4,700	47.2	90.5	2,620	36.1	90.5
2,500	4,730	20.9	124.5	3,020	17.2	124.5
3,333	4,830	8.03	124.7	3,680	7.54	124.7
4,000	4,820	4.08	104.1	4,480	3.96	104.1
4,250	4,860	3.37	95.7	4,860	3.37	95.7
5,000	4,750	1.95	72.0	6,220	2.02	72.0
6,667	4,070	0.61	37.5	9,780	0.66	37.5
10,000	2,450	0.069	11.7	16,200	0.079	11.7

The most convenient form in which to express the character of the interstellar radiation, when it is to be used in calculating ionization, is in terms of a number of simultaneously acting sources of black body radiation, each diluted by a different factor. For the case where no absorption was assumed this is a simple matter because the data were in precisely that form before they were summed to give the combined curve. For the case with absorption the combined curve was resolved as satisfactorily as possible into several black body curves with suitable dilution factors. The resulting parameters are given in Table III. It is interesting to notice that the total radiation density for the case without absorption, in the fifth column, adds up to 5.24×10^{-13} ergs/cm^3, which is in good agreement with the value 7.67×10^{-13} ergs/cm^3, which was used by Eddington.[24] The values for λ10,000 in the case with absorption are unreliable and high owing to the large positive corrections for selective absorption at this wave-length. This region of the spectrum has no significant effect on ionization and might have been omitted. (The large correction in the infra-red is responsible for the anomalous result that the total density of radiation is greater with absorption than without.)

When multiple sources, each diluted, are acting simultaneously to produce ionization, the degree of ionization is given by

$$\frac{N_1}{N_0} = \frac{X}{1-X} = \left(\frac{x}{1-x}\right)_a + \left(\frac{x}{1-x}\right)_b + \left(\frac{x}{1-x}\right)_c + \cdots,$$

TABLE III

DILUTE BLACK BODY SOURCES REPRESENTING INTERSTELLAR RADIATION

T	Type	Equiv No. Stars 0 0 m_{pg}	Without Absorption			With Absorption		
			$\delta\times10^{16}$	Ergs $\times10^{13}$	% Energy	$\delta\times10^{16}$	Ergs $\times10^{13}$	% Energy
65,000	W–O5	.212	.00071	.097	1.8	.0000107	.00145	.012
49,000	O6–7	.288	.00159	.070	1.3	.00024	.00105	.008
33,000	O7–8	.253	.00256	.023	0.4	. .	. .	. .
23,000	B0–2	8.39	.158	.338	6.4	.0788	.169	1.36
16,000	B3–9	28.58	1.17	.586	11.2	.293	.147	1.17
10,000	A0–9	56.16	8.99	.687	13.1	13.5	1.03	8.25
6,400	d(F0–G9)	89.01	97.8	1.255	24.0	. .	. .	. .
5,000	g(F0–G9)	32.06	122.8	.586	11.2	. .	. .	. .
4,400	d(K0–M)	5.54	67.1	.192	3.7	. .	. .	. .
3,700	g(K0–M)	18.92	985.	1.410	26.9	7780.	11.14	89.20
				5.244	100.0		12.49	100.00

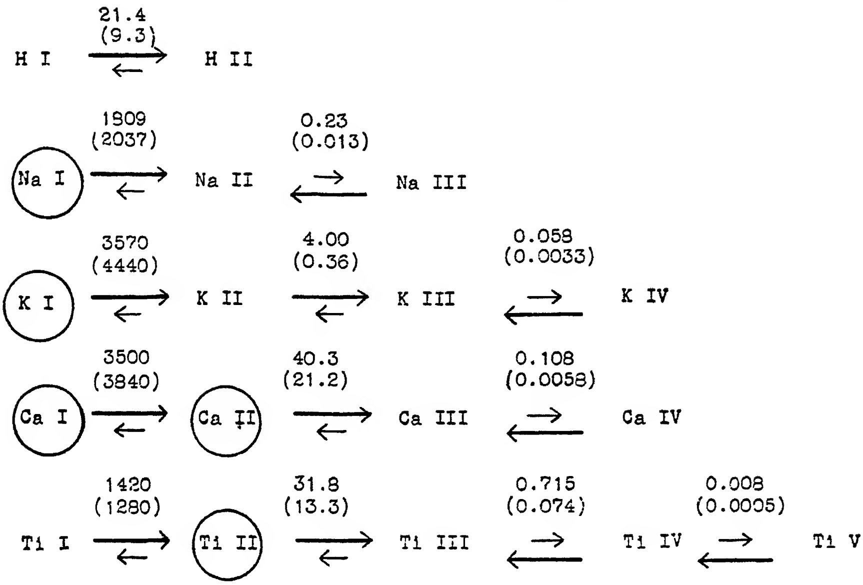
where

$$\left(\frac{x}{1-x}\right)_n = \frac{-5040I}{T_n} + \log T_n + \tfrac{1}{2}\log T_e + \log\frac{u_{r+1}+2}{u_r} - \log N_e + 15.38 + \log\delta.$$

x = fraction of atoms ionized,
I = ionization potential,
T_n = effective T of nth source,
T_e = temperature of electrons (see Pannekoek [25]),
u_{r+1} = statistical weight of more highly ionized state,
u_r = statistical weight of less highly ionized state,
N_e = number of electrons per cm^3,
δ = dilution factor.

On introducing the value, 3500, which has been found for *Ca* II : *Ca* I we obtain $N_e = 14.4$ if there is no absorption, and $N_e = 7.30$ if there exists the amount of general and selective absorption which was previously assumed. By using these values for the electron density, all of the other ionization ratios with which the present data are concerned may be calculated. The results are shown in Fig. 3, where each number refers to the ratio of the concentration of the ion following to that of the ion preceding. The numbers in parentheses are the ratios when absorption is taken into account. Observed ions are shown enclosed in circles.

FIG. 3. Ionization equilibria



The ionization ratios shown in Fig. 3 have been applied to the concentration of the observed ion in the case of each element in order to obtain the concentrations of all the ions. These are given in Table IV. It seems likely that the ionization of *Ti* II is blocked by the Lyman

TABLE IV

PARTICLES IN ONE CUBIC METER OF INTERSTELLAR SPACE BETWEEN THE EARTH AND χ^2 ORIONIS

Stage of Ionization	Electrons	Hydrogen	Sodium	Potassium	Calcium	Titanium
I	14,400,000 (7,300,000)	700,000 (800,000)	0.050 (0.050)	0.00083 (0.00083)	0.00004 (0.00004)	0.000,00091 (0.000,0010)
II		14,400,000 (7,300,000)	90.0 (102)	2.96 (3.68)	0.14 (0.14)	0.0013 (0.0013)
III			21.0 (1.34)	11.8 (1.33)	5 65 (2.97)	0.042 (0.017)
IV				0.69 (0.004)	0.61 (0.017)	0.030 (0.0013)
Total Without abs. With abs.	 14,400,000 7,300,000	 15,100,000 8,100,000	 110. 103.	 15.5 5.0	 6.40 3.14	 0.073 0.020

$$\text{Ratio } \frac{Na}{Ca} = 17, \text{ without absorption,}$$
$$= 33, \text{ with absorption.}$$

continuous absorption of hydrogen since the head of the *Ti* II series is at λ909, just 3A beyond the hydrogen limit. If this is so, the total amount of titanium will be greatly reduced, since there will be little contribution by *Ti* III and *Ti* IV.

The ionization equation, as used here and by other workers previously, does not apply strictly to the conditions of interstellar space. The recombination processes between ions and free electrons proceed exactly as in thermodynamic equilibrium, but the ionization processes are markedly reduced, because ionization can occur only from the ground state. The work of Cillié indicates that the ionization of hydrogen will be reduced by a factor close to two. In the case of *Na* I, Dr. H. Zanstra, to whom I am indebted for valuable advice on this subject, estimates that the factor may be ten or more. The ratios for the heavier ions in Fig. 3 would not be affected if the correction factor were the same for all ions, since the value of N_e is itself derived from a calculation subject to this correction. It is probable, however, that N_e, and hence the abundance of *H*, should be reduced by a factor of ten or more, and this would lead to a hydrogen abundance which might be somewhat less than one atom per cm^3.

The relative abundance of *Na* and *Ca* has been the basis of extensive discussion by many writers. The ratio derived in the present work is very different from the value (close to unity) which applies in the earth's crust and in the solar atmosphere. Perhaps it should not disturb us if the relative abundance of the elements is somewhat different in interstellar space.

It should be borne in mind that the present results are based on a single sounding of space in the direction of one particular star. The concentration of *Na* I appears to be considerably higher than the value (0.003 atom per cubic meter) obtained by Wilson and Merrill,[21] which was an average for all directions. Evidently local clouds are very pronounced and it may well be that such a cloud between the earth and χ^2 Orionis is responsible for the detection of the neutral calcium line in this star. A representative census of interstellar atoms must be based on much more extensive observational data.

The black body intensities used in this investigation were taken from an unpublished table computed by Miss Dorothy Carlson, who has helped in the compilation of much of the statistical data required in estimating the quality of interstellar radiation.

REFERENCES

1. HARTMANN, *Ap. J.*, **19**, 268, 1904.
2. MISS HEGER, *Lick Obs. Bull.*, **10**, 141, 1919.
3. DUNHAM, *Phys. Rev.*, **46**, 326, 1934.
4. DUNHAM AND ADAMS, *Publ. Am. Astr. Soc.*, **9**, 5, 1937.
5. DUNHAM, *Nature*, **139**, 246, 1937.
6. DUNHAM, *Publ. Astr. Soc. Pac.*, **49**, 26, 1937.
7. MERRILL, *Publ. Astr. Soc. Pac.*, **46**, 206, 1934.
8. SWINGS, *Monthly Notices R. A. S.*, **97**, 212, 1937.
9. MERRILL, *Phys. Rev.*, **52**, 761, 1937.
10. MERRILL, SANFORD, WILSON AND BURWELL, *Ap. J.*, **86**, 274, 1937.
11. DUNHAM, *Publ. Am. Astr. Soc.*, **7**, 215, 1933.
12. ALLEN, *Commonwealth Solar Obs. Memoir*, **5**, Pt. 1, 21, 1934.
13. WILLIAMS, *Ap. J.*, **83**, 305, 1936.
14. MENZEL, *Ap. J.*, **84**, 462, 1936.
15. EDDINGTON, *Internal Constitution of the Stars*, p. 383, 1926.
16. GERASIMOVIČ AND STRUVE, *Ap. J.*, **69**, 19, 1929.
17. Carnegie Inst. of Wash., Year Book, No. **37**, 176, 1938.
18. LANGER, *Bull. Amer. Phys. Soc.*, **13**, Nos. 6, 7, 1938; and *Phys. Rev.*, **55**, 423, 1939.
19. STRUVE AND ELVEY, *Ap. J.*, **88**, 364, 1938.
20. STRUVE, *Proc. Nat. Acad. Sci.*, **25**, 36, 1939.
21. WILSON AND MERRILL, *Ap. J.*, **86**, 44, 1937.
22. FILIPPOV AND PROKOFJEW, *Zeits. f. Phys.*, **56**, 458, 1929.
23. UNSÖLD, *Physik der Sternatmosphären*, p. 197, 1938.
24. EDDINGTON, *Internal Constitution of the Stars*, p. 371, 1926.
25. PANNEKOEK, *B. A. N.*, **3**, 207, 1926.

STELLAR ENERGY

HENRY NORRIS RUSSELL

Professor of Astronomy and Director, Princeton University Observatory

(Read on February 17, 1939, in Symposium on Progress in Astrophysics)

THIS lecture should begin by an apology to the audience. The speaker tonight should be the man whose recent and brilliant work has inaugurated a new and very promising stage of astrophysical study—Professor H. A. Bethe of Cornell. The planners of the program can plead but one excuse—his results had not been announced when it was prepared.

They are not yet published in detail; but Professor Bethe has most generously given me a full copy of the proof of his forthcoming paper [1], and amplified this in personal conversation. What there is of novelty in the present discussion should be credited entirely to him, and not to myself. Indeed my report, apart from this, is mainly upon the work of others in a field in which I did a little pioneering a good many years ago.

A century of precise measurement has furnished astronomers with a large accumulated capital of facts about the stars. One of the most remarkable things that is thus revealed is that the stars differ greatly among themselves in some properties, and relatively little in others. For example, the luminosity—the rate of radiation of energy—ranges (roughly) from a million times the Sun's luminosity to a thousandth part of it, a ratio of a billion to one. The diameters run from at least 1000 times the Sun's down to 1/30 or less, a range of thirty-thousand-fold. For the masses, the range is roughly from 100 times the Sun to 1/10 of the Sun, a thousand-fold. But surface temperatures more than ten times the Sun's or less than 1/4 that of the Sun are practically unheard-of, so that the known range is only forty-fold.

Why should such enormous diversities exist? The limitation of our study is set, on one side, by the obvious fact that we cannot see a star unless it shines—that is, unless it gives out enough light to be perceptible at stellar distances. This obviously sets a limit to our knowledge of *faint* stars, and of *small* stars—which, other things being equal, will shine less brightly because they have fewer square miles of surface, and

[1] *Phys. Rev. 55*, 434, 1939.

still more to the study of *cool* stars (of low surface temperature). The heat radiation per square mile varies as the fourth power of the temperature, and a forty-fold increase in the latter changes the former by a factor of about two and a half million. Moreover, at low temperatures, most of the radiated energy is in long infra-red waves, invisible to the eye, so that a star—like a mass of hot iron—ceases to be visible in the dark, long before it stops giving out heat.

But there is no such limitation on the side of great brightness. If we fail to find still brighter stars—though we are fishing for these in very wide waters—it must be because there are no such fish in our sea. Nature herself must in some way set a limit to the brightness attainable by a star—at least as a permanent affair, for we know that the short-lived outbursts of supernovæ are enormously more intense. Such a limitation would explain why stars of very large diameter are never found to be very hot—if they were, they would give out too much light; but it has no obvious relation to the mass of a star.

A notable advance was made more than twenty years ago when it was found observationally (first by Halm) that there is a close relation between the mass of a star and its luminosity. All later work has confirmed this, and it now appears that, over almost if not quite the whole available range, the total heat-radiation of a star is nearly proportional to the fourth power of its mass. The most remarkable feature is that, for a given mass, the luminosity does not depend much upon the size of the star—the large ones are cool, and give out less heat per square mile, the small ones hot, and give out more; but the net product is nearly the same.

Eddington's explanation of this, completed fifteen years ago, marks the first, and, even now, the greatest success of modern atomic physics in interpreting the stars.

Since the work of Lane in 1870, it has been realized that, if the familiar laws of perfect gases could be applied to the interior of the Sun, the central temperature must be many millions of degrees. The result may be written

$$T = 11.5 \frac{M}{R} \frac{Y}{X} \mu,$$

where T is the temperature in millions of degrees, M the star's mass and R its radius, μ the mean molecular weight of the material, and Y and X quantities depending on the "density-model," that is, the law in accordance with which the density increases toward the center.

Calculations for different laws of density-distribution show that the "model" has a rather small influence. If the density were everywhere

the same, X and Y would be 1 at the star's center. On the model to which Eddington's calculations led, the central density was 54 times the average ($X = 54$); but $Y = 92$, so that the central temperature for the Sun comes out 19,700,000°. A model recently calculated by Chandrasekhar makes $X = 88$, $Y = 171$, and the temperature 22,300,000°; still another, reported by Bethe, 20,300,000°—all provided that $\mu = 1$. There are good reasons to believe that these "models" give a fairly correct idea of the situation, and therefore that the central temperature of the Sun is close to twenty million degrees—provided that the molecular weight is 1.

But, under ordinary circumstances, μ is 1 for hydrogen, 4 for helium, 16 for oxygen, 56 for iron, and so on, so that the calculated internal temperature of the Sun would depend enormously on its assumed chemical composition.

Fortunately for us theorists, the situation is simpler inside the stars; for the outsides of the atoms—the electrons which surround the nucleus—are pretty well knocked off. A hydrogen atom (with one electron) is broken into 2 particles, a helium atom into 3, one of oxygen into 9, and one of iron into 27 (or perhaps 25, if the two tightest bound electrons stay with the nucleus).

The average mass of a free particle (which is what counts) then comes out 1/2 for hydrogen, 4/3 for helium and nearly 2 for all heavier atoms. So we can conclude that the central temperature of the Sun is close to 10 million degrees if it is pure hydrogen, 27 million if it is all helium, and 40 million if it is all composed of heavy atoms—and the result for any assumed mixture can easily be computed.

The temperature of the Sun's surface is but a few thousand degrees. Heat will flow outward from the hot interior "down grade" and the surface will adjust its temperature so that it loses by radiation into space just the amount which it gains by transmission from the deep interior. The rate of this transmission thus determines how bright and hot the Sun (or any other star) will be; it takes place mainly by the transfer of radiation from atom to atom through the gas; and the net opacity, which determines the rate of transfer, may be calculated with considerable accuracy, on principles developed by Kramers.

When the net flow of heat out to a star's surface is thus calculated, it is found that it depends relatively little on the "model," and not much upon the radius, but changes rapidly with the mass—at very nearly the rate indicated by observation.[1]

The mass-luminosity relation is thus explained—and with it the

[1] Certain complications which arise in the more massive stars, where radiation-pressure becomes a sensible fraction of gas-pressure, are neglected in the above account, for brevity.

fact that stars of mass less than one-tenth the Sun's have not so far been detected—they are presumably too faint to be seen; and our failure to discover very large masses is connected with the absence of stars of very great luminosity.

Eddington's theory, however, predicts that the luminosity of a star, other things being equal, should change very rapidly with the mean molecular weight of its material (about as its seventh power). A star composed almost entirely of heavy atoms (with $\mu = 2$) will be so hot inside that it will shine brightly, whereas one composed mostly of hydrogen will be cooler inside—and fainter outside by several hundred-fold (the exact amount involving refinements of calculation omitted here).

The observed luminosity of the Sun is about midway between these limits, and may be accounted for, according to Strömgren's calculations, if one third of the material, by weight, is hydrogen, and the rest heavy atoms; or with more hydrogen if helium is present; for example, 60 per cent hydrogen, 36 per cent helium, and 4 per cent heavy atoms. The mean molecular weight comes out 0.98 and 0.67 in the two cases, and the central temperature 19 million and 13 million degrees.

Repeating his calculations for other stars for which we have reliable data, Strömgren finds very nearly the same composition for Sirius A—the bright star, not its faint companion—for Capella, and for several others. Some stars contain less hydrogen, *e.g.* ζ Herculis A, which is four times as bright as the Sun though of very nearly the same temperature and mass. Its computed hydrogen content is 12 per cent (if there is no helium).

The observed mass-luminosity relation is now completely accounted for on principles derived from general atomic theory—but a new phenomenon takes its place to be explained. Why should so many of the stars contain almost the same amount of hydrogen? This is obviously not the kind of question which can be answered by a general theory of the type already described; and there are other much more obvious properties of the stars of which this is also true. For example, if we plot the luminosities (or the stellar magnitudes) of the stars against the surface temperatures (or the spectral types) we obtain a now very familiar diagram which shows marked peculiarities. The greater portion of the dots representing the stars lie in a narrow strip—indicating the existence of a "main sequence," having at one end bright, white stars, of great luminosity, large mass, and high surface temperature; and at the other faint red stars, their opposites in all three respects. The Sun is a typical member of this sequence, about midway in it. Calculating the diameters, we find that the stars at the top of the se-

quence are larger and less dense than the Sun, those at the bottom smaller and denser; but these differences are not very great, and most of the enormous range in brightness from one end to the other arises from the differences in surface temperature. The central temperature varies still less—it is some 30 or 35 million degrees for the stars at the top of the list, and may be as low as 12 or 15 million for the faint red ones at the bottom.

Almost as conspicuous as the main sequence (and indeed more so in a list of naked eye stars, which automatically picks the bright ones) is a wider, but definite, band, beginning with stars of about the Sun's surface temperature (Class G) but some fifty times the Sun's brightness, and extending to some of the reddest known stars, which give off several thousand times as much heat (though not visible light) as the Sun.

These giant stars (so named for their brightness) deserve the name for their size. Capella, which is a little cooler than the Sun, and 120 times as luminous, must have about 16 times the Sun's diameter. At the other end of the giant sequence come huge objects several hundred times as big as the Sun. The masses—according to the rather scanty evidence which alone is available, after careful searching—are about the same as for main-sequence stars of equal luminosity. For Capella alone is an accurate value known, 4.2 times the Sun's mass. This makes its density hardly more than 1/1000 that of the Sun, and its central temperature (if built on a similar model) less than 6 million degrees. For the great red star ζ Aurigæ, the mean density is one nine-millionth part of the Sun's, and the central temperature (for the same model) but 1,200,000 degrees.

At the top of the diagram are scattered points representing the supergiant stars, extremely luminous, and with all sorts of surface temperatures. They must be of large diameter, and low density and internal temperature. The region below the giants and to the right of the main sequence is almost vacant, as is also that below the main sequence to the left. It is in this last region that we should find points representing stars with internal temperatures higher than twenty or thirty million degrees—and there are hardly any[1]—except down at the bottom, the famous white dwarfs (like the companion of Sirius), somewhat less massive than the Sun, less than one per cent as bright, and comparable in diameter with Uranus or Neptune. The mean densities of these bodies are as great as 100,000 times that of water, and they represent a "degenerate" state of matter, in which as many electrons are jammed into every cubic centimeter as the quantum-laws will permit. They

[1] There are a few white stars about as bright as the Sun, for which we have not yet sufficient data to say anything definite.

are surprisingly well understood, but do not belong to our present problem.

The luminosity-spectrum diagram reveals one more important fact. There are a great many stars with central temperatures of 20 to 30 million degrees, a good many with lower central temperatures, and very few, if any, which are hotter inside, except perhaps the white dwarfs. Why should this be? This ties up with an older and more fundamental question:—Why do the stars keep on shining?

When first the Sun's heat was measured, it was realized at once that it could hardly be maintained, even for the duration of history, by ordinary chemical combinations—what we now describe as reactions between whole atoms. A far larger source of energy is found in a slow contraction, compressing the gas and turning the gravitational potential energy into heat, and this explanation, due to Helmholtz and Kelvin, was still accepted forty years ago. The energy-supply thus available for the Sun's past history is about thirty million times its present annual expenditure (depending somewhat on the assumed model). But at least half this must still be stored inside the Sun in the form of heat, so that less than fifteen million years of sunshine can be accounted for.

But the Sun has been shining and warming the Earth, very much as now, throughout geological history, which is certainly a hundred times as long as this. There is only one place in the known universe to look for so vast a story of energy—and that is in the minute nuclei of atoms. From these nuclei is liberated the energy of radio-activity—which is of the required order of magnitude, though rather small. But radio-activity itself will not meet our needs, for it takes place at a rate uninfluenced by conditions external to the nucleus—such as the surrounding temperature. To account for the observed variation of luminosity with mass, we would have to assume that in some inscrutable way the amount of active material had been exactly proportional "in the beginning" to the mass of each star, so as to provide the proper heat supply to maintain its radiation; and also that practically none of it had been allowed to enter the Earth or the other planets—for otherwise their surfaces would be red-hot or even white-hot.

It is necessary, therefore, to assume that the great store of energy upon which the stars draw becomes available only under stellar conditions—that is, obviously, at temperatures of millions of degrees.

The whole course of nature teaches that it would not become available abruptly, but at a rate increasing with the temperature, and probably increasing very rapidly. But here appears a difficulty. If too much heat should at any time be liberated inside a star for the

steady process of escape to the surface to carry off, would not the material grow hotter, "turn on" a still more rapid supply of heat, till some tremendous explosion ensued?

The answer lies in a curious property of a gravitating star. As we have already seen, it must grow hotter inside if it contracts; but the gravitational energy released by the contraction is more than enough to supply this heat, and leaves a balance available for radiation. Only as heat is lost into space can the temperature of the interior rise. If just enough heat is liberated from some sub-atomic process, there will be a steady state; if too little, contraction will supply the rest; if too much, this contraction will be converted into an expansion. This expansion will lower the internal temperature, and shut off the over-supply. The Sun has an internal store of heat equal to at least eight million years' outgo. The sudden introduction into the interior of even a thousand years' heat supply would therefore produce very little effect (unless it were localized in a small region, when it might cause trouble).

One other thing was clearly understood, twenty years or so ago; the reactions which provide the stars with their energy must be accompanied by a perceptible loss of mass. The theory of general relativity indicated that mass and energy should be interconvertible, at the rate of c^2 units of energy (ergs) for one unit of mass (gram), where c is the velocity of light. The Sun's rate of loss of energy by radiation (3.8×10^{33} ergs per second) is far too great to be comprehensible in ordinary terms; but, if measured in mass-units, it means that the Sun is actually losing 4,200,000 *tons* of energy every second! The Sun's mass (2×10^{33} g) is so great that even at this enormous rate, it would take 15 billion years to reduce it by one part in a thousand.

The nature of the process—or, at least, of one process—by which this transformation of mass into energy can take place, is now well understood, since it can be made to happen in the laboratory, in many different ways. This is the transformation of atoms of one sort into another, with the appearance of an amount of energy corresponding to the change in total mass; and it is to this that we now look as the source of stellar energy.

The alternative hypothesis, that positively and negatively charged particles (protons and electrons) may utterly annihilate one another, with the appearance of the whole amount of energy corresponding to the sum of their masses, has now been dropped, since no evidence of it has been observed. It does appear to happen for positive and negative electrons; and judgment may well be reserved about "heavy electrons" (mesotrons) till more is known about them.

The masses of the lighter atoms (including their outer electrons) are given in Table I. They can be very accurately determined—mainly

TABLE I

MASSES OF LIGHT ATOMS [1]

Name	Symbol	Charge	Mass
Electron	e	−1	0.00055
Neutron	n	0	1.00893
Hydrogen	H	1	1.00813
Deuterium	H^2	1	2.01473
Helium	He^3	2	3.01699
"	He^4	2	4.00386
Lithium	Li^6	3	6.01686
"	Li^7	3	7.01818
Beryllium	Be^9	4	9.01504
Boron	B^{10}	5	10.01631
"	B^{11}	5	11.01292
Carbon	C^{12}	6	12.00398
"	C^{13}	6	13.00761
Nitrogen	N^{14}	7	14.00750
"	N^{15}	7	15.00489
Oxygen	O^{16}	8	16.00000

[1] These values are from Bethe's computations, with some changes from his latest work.

with the mass-spectrograph. The various isotopes—atoms of different weight, but the same chemical properties—are of course listed individually.

From the table it appears that if four hydrogen atoms could in any way be converted into one helium atom, a loss of mass of 0.02866 unit would result. This is 1/141 of the original mass. If, therefore, a thousandth part of the Sun's mass were hydrogen, and could be transmuted into helium, energy enough would be released to keep the Sun shining for 106 million years.

The excess of the tabulated mass above an even value evidently represents the possible value of the given atom as a source of energy. This excess is fairly considerable for most of the lighter atoms; but oxygen, carbon and helium are relatively very stable. Some atoms heavier than oxygen are still further "down"; but the differences are small and need not concern us here. Per unit of mass, hydrogen is much the best source of energy.

Two types of nuclear reaction, in the laboratory, give rise to reactions of the sort considered:—the penetration into a heavier nucleus of a neutron, or of a charged particle (proton, deuteron, or alpha-particle). The former are the easiest to produce artificially—since the other nucleus does not repel the neutron. They occur in great variety, and

liberate large amounts of energy. For this very reason, it appears that they are not of much astrophysical importance. They happen *too* easily. If there were a lot of neutrons in the interior of a star, they would all collide with atomic nuclei of one sort or another—sometimes building up a heavier atom, sometimes causing the emission of some other particle—and they would be used up in the process, in the twinkling of an eye. There might be an almost explosive outburst of heat (under the altogether unnatural conditions which we have imagined) but anyhow the neutrons would be gone, and similar reactions would occur in future only if "new" neutrons were produced by other nuclear reactions. A careful study by several investigators has shown that this should happen so very rarely that its effects can be neglected.

We are left, then, as a steady source of energy, with the penetration of one charged nucleus into another. The repulsion between the two is always so great that only an exceptionally violent head-on collision between particles moving much faster than the average would be effective. The probability of such a collision has been calculated. It diminishes very rapidly with the charges of the particles, so that a proton is vastly more likely to succeed than an alpha-particle, while, even so, it stands very little chance of getting into a nucleus with a larger charge than oxygen (provided, at least, that the temperature is not more than twenty million degrees).

What happens after the proton gets in depends on the individual peculiarities of the nucleus which it has hit. It may simply go in and stay—producing a new nucleus greater by 1 in both charge and mass; or a positive electron may be ejected, leaving a nucleus of the same charge as before but of greater mass; or the old nucleus may break up into two or more pieces—one of which is usually an alpha-particle. None of these changes can happen if the hypothetically resulting nucleus is heavier than the reacting particles, for a quite impossible amount of energy would have to be supplied from no known source. If the mass of the product is less than that of the two particles, the excess energy will appear as kinetic energy when the nucleus breaks up, or an electron is ejected;[1] in the first case, it appears as radiation—a gamma-ray.

All three processes occur in the stars.

The astrophysical importance of reactions of this type was first pointed out by Atkinson. They have been discussed by von Weizsäcker, Gamow and Bethe, and enough is now known about the properties of nuclei to give a pretty clear picture of what would happen.

[1] Much of the energy in this case is believed to be carried off by that elusive particle, a neutrino.

First. The one stable thing in this microcosm of change is the alpha-particle (He^4). Apparently nothing can happen to it. The nuclei which might imaginably be formed by collision with a proton (He^5 or Li^5) are unstable, and do not exist at all; a collision of two alpha-particles would give Be^8, which is slightly unstable, and breaks up again. Three fast-moving alpha-particles, colliding simultaneously, might possibly form C^{12} with liberation of energy; but such an event is so very improbable that it can be neglected, except at much higher temperatures.

If, then, hydrogen is "the fuel of the stars" helium is the ashes.

Second. The other light nuclei, up to and including boron, are highly susceptible to proton collisions. The different ones go through various transformations, but every sequence ends irrevocably in helium —for example, $Li^7 + H = 2\ He^4$, $B^{11} + H = 3\ He^4$. (In this last case the nucleus breaks into three pieces—all helium.) Hence these elements, if originally present in a star, would be successively exhausted as the reactions went on. If we imagine a mass equal to the Sun's slowly contracting from a large size, with increasing central temperature, then, according to Bethe, the reaction which eats up deuterium would happen fast enough to supply the star's radiation, when the central temperature was 360,000° (more or less, depending on the amount present).

So long as any considerable amount of deuterium remained, the star would not contract further. As it was exhausted a new contraction would begin, to be halted when the central temperature was about 2,000,000° and lithium began to be transformed and used up. Beryllium would have a similar fate at a central temperature of 3½ million degrees, and the two isotopes of boron at five and nine million. Since the Sun's central temperature is much higher than this, there can be practically none of any of those elements left in its interior. At the surface, spectroscopic evidence shows that all three are present, but in very small quantities.

Third. At a temperature somewhat above fifteen million degrees, carbon begins to be attacked, and something quite new happens, which is tabulated by Bethe as follows:

$C^{12} + H$	$= N^{13} + \gamma$	2,000,000 years
N^{13}	$= C^{13} + \epsilon^+$	10 minutes
$C^{13} + H$	$= N^{14} + \gamma$	50,000 years
$N^{14} + H$	$= O^{15} + \gamma$	50,000,000 years
O^{15}	$= N^{15} + \epsilon^+$	3 minutes
$N^{15} + H$	$= C^{12} + He^4$	2,000 years.

The proton goes into C^{12} and builds up N^{13} (the spare energy escaping as a γ-ray). This is one of the artificial radio-active elements, which has been produced in the laboratory. It emits a positive electron (ϵ^+) and goes over into the carbon isotope C^{13}. The next proton turns this into ordinary nitrogen N^{14}, and another produces a radio-active oxygen O^{15} which goes over into "heavy" nitrogen N^{15}. A proton collision might build this up into ordinary oxygen O^{16}; but Bethe calculates that it is ten thousand times more probable that it would split the nucleus into two parts, one of which is helium, and the other the original carbon C^{12}, ready to be used over again! The carbon is not consumed, but acts as a catalyst for the transformation of hydrogen into helium, and we have a regenerative process, which should continue until all the hydrogen has been transformed.

Four of the six steps here described have already been observed in the laboratory. The other two (involving collisions of protons with nitrogen) can reasonably be inferred from known data. The average length of time during which an atom of each kind may be expected to last before its next adventure, is given in the table. The $N^{14} + H$ reaction is the slowest; it will then be the "bottle-neck" for the whole process. Bethe's calculations show that it would supply energy enough for the Sun at a temperature of 18,300,000° (assuming 35 per cent of hydrogen and 10 per cent of nitrogen). As the rate of reaction increases as the 18th power of the temperature, it would do the work, with 1 per cent of nitrogen, at a temperature of 20,800,000°. Now these temperatures, which are calculated from pure nuclear theory, agree excellently with the central temperature of the Sun, as calculated from astrophysical data. An almost equally good agreement is found for Sirius (22 million from theory, 26 from observation), and for the very hot and massive star Y Cygni (30 and 32 million). Moreover, the gradual changes in light, diameter, density and surface temperature for stars of different mass along the whole main sequence are in excellent agreement with the results of the theory.

It appears, then, that one great part of the problem of the source of stellar energy has been fully solved. The main sequence is explained in detail.

There are still worlds left to conquer. The very rapid energy-liberation in the giants cannot be accounted for in the same way. It might be assumed that they are built on the same sort of model as the main-sequence stars, but contain larger quantities of lithium, beryllium or boron—but this is a far-fetched idea, and does not explain why they form a sequence in which the reddest stars are the brightest. It is also

possible, as Gamow, Öpik, and others have suggested, that they are built on quite a different model, with relatively small and dense cores, which may be as hot as those of the main-sequence stars or hotter. Much more is likely to be done on this problem within a few years.

Reactions involving heavier nuclei are possible, but they would occur at significant rates only at higher temperatures, and so would not have a chance to happen if there were any carbon in the stars—as there assuredly is. If, at the other extreme, we had a star composed originally of pure hydrogen, it might get energy by the following chain of reactions, also according to Bethe:

$$\begin{aligned} H + H &= H^2 + \epsilon^+ \\ H^2 + H &= He^3 \\ He^3 + He^4 &= Be^7 \\ Be^7 + \epsilon^- &= Li^7 \\ Li^7 + H &= 2\,He^4. \end{aligned}$$

Two protons form a deuteron—a slow reaction; the latter goes over almost at once into He^3; this with an alpha-particle forms a known isotope of beryllium, which turns to lithium and then to helium—the net result being, as usual, the building of four hydrogen atoms into one of helium.

The rate of this process changes much more slowly with the temperature than that of the carbon cycle. Bethe suggests that it may supply the energy in the coolest stars of the main sequence. Present data do not suffice for a decision.

Two large questions remain to be mentioned.

First. What will be the history of a star? The first portion, till the carbon-cycle is "turned on," has already been described. The star will then keep shining as a member of the main sequence, for a very long time—gradually using up its hydrogen, increasing the mean molecular weight, and growing brighter, hotter, and a little larger. At last, when the hydrogen is exhausted, it must contract, become fainter as well as smaller, and end up, if of small mass, as a white dwarf. A large mass should contract to a still higher density—how far, no one can yet say. The Sun should have a good ten billion years to go before it begins to fail. Meanwhile as it grows more luminous, the Earth should gradually grow warmer—at the rate of about 1° Fahrenheit in 100 million years.

Finally, Where did the carbon, in the Sun—and also the oxygen and heavier atoms—come from? There is nothing in the reactions sketched above which can form them, and the careful studies of von Weizsäcker and Bethe have found no chance of the production of more than the

merest traces by any side-reactions. Yet present evidence indicates that nearly two-thirds of the Sun is composed of such heavy atoms—for, if there were fewer of them, and more helium, the molecular weight would be less, and the internal temperature too low to start the carbon-chain running.

It seems necessary to assume that they are *older than the stars*—constituents of the primitive matter of which these were formed. Von Weizsäcker has suggested—in accordance with certain forms of the theory of the expanding universe—that, at some remote time, the matter in the universe, or a great part of it, was concentrated into a vastly smaller compass than at present, and was much hotter than the inside of the hottest stars are now; but this is highly speculative, though highly interesting.

Another and more practical riddle is: What have huge stars like Y Cygni been doing throughout the 2000 million years or so since the universe fairly started to expand? Such a star is spending its store of energy at least five hundred times faster, in proportion to the total supply, than is the Sun. The transformation of the whole mass from hydrogen to helium would keep it going, at its present rate, for only about 200 million years. Whether this means that these massive stars have begun to shine rather late in the history of the Galaxy, or that, after all, they have still greater stores of energy to draw on, in ways as yet unknown, must be left to the future to determine. But it is reasonable to suppose that, if still more massive stars, which would be far brighter than any now known, had once existed, they would have burned themselves out by this time—so that the question with which we started can be at least tentatively answered.

ON THE EVOLUTION AND MAJOR CLASSIFICATION OF THE CIVETS (VIVERRIDAE) AND ALLIED FOSSIL AND RECENT CARNIVORA: A PHYLOGENETIC STUDY OF THE SKULL AND DENTITION

WILLIAM K. GREGORY AND MILO HELLMAN

Contents

Abstract

We find that the terms Creodonta, Fissipedia and Pinnipedia are somewhat arbitrary and Matthew's divisions more natural. Matthew's Eucreodi might well include both Fissipedia and Pinnipedia, as evidence shows that all families of modern Carnivora were derived from one ancestral stock. A new infraorder, the Amphicreodi, is here erected to include the transitional group of Eocene Palaeonictidae. For reasons stated, we regard Matthew's Miacidae as forerunners of later civets and accordingly we include this family with the Viverridae. We regard Pocock's families and subfamilies as "natural groups" but believe that his arbitrary choice of diagnostic characters separates many genera from their true relationships and have therefore grouped his subfamilies into sections, using the termination *ida* to designate them. We consider it probable that the most primitive types of cheek teeth are those illustrated by the Paleocene *Propappus* and *Didymictis* and to a less extent by the modern *Viverra* and *Herpestes*. We have found the auditory bullae in some cases more conservative than the dentition and valuable as diagnostic characters, at least for subfamilies. A summary of the evolution and adaptive radiation of the auditory bullae in the viverrids and their allies is given here.

In general our studies incline us to question current applications of the doctrine of irreversibility of evolution to the major classification and phylogeny of the Carnivora, and to suggest that among the aeluroid Carnivora, as well as in all other groups, earlier specializations are often wiped out or transformed into later specializations. The reduction and loss of parts have accompanied the emphasis of other parts. The systematists' habit of "splitting" tends to obscure the earlier connections between supergeneric groups.

Introduction

Some years ago in the course of our studies on the evolution of the dentition, jaws and skulls of Primates, we began to review the evolution of these parts among the fossil and recent Carnivora as a basis for contrast and comparison with the Primates. But as we were convinced

by experience that morphological study can seldom lead to sound results if the evolution of particular parts is considered while the evolution of the animals that bore those parts is ignored, we soon found ourselves in the midst of a multitude of phylogenetic-morphogenetic problems involving such questions as the following:

(1) Considering each of the families and subfamilies of fossil and recent Carnivora, how may we describe and measure the differences in the general appearance and in the component parts of the skull, jaws and dentition of the included genera?

(2) Which genus in each subfamily appears to have the most central and primitive conditions of the skull, jaws and dentition? By what steps have the specialized extremes been attained?

In a previous paper [1] one of us has dealt with the evolution of the jaws and dentition of the arctoid group. The present paper is concerned with the numerous and diverse beasts that are assembled in the Aeluroidea of Flower. Of this series the principal forms are as follows: (1) the perfume-bearing civets (*Viverra*) of Asia, the genets (*Genetta*) of Europe and Africa, the African civet (*Civetictis*), together with certain less widely known genera, *Osbornictis* and *Poiana* of Africa; (2) the night-prowling and strong-jawed hyaenas and their degenerate relative, the South African aardwolf (*Proteles*); (3) the genet-like linsang (*Prionodon*) of India, which has shearing carnassial teeth and seems to point toward (4) the cat-like fossa (*Cryptoprocta*) of Madagascar; this in turn tends to connect (5) the entire Cat family (Felidae) with the viverrine stem; (6) the arboreal palm civet and toddy cat (*Paradoxurus*) and the "bear-cat," binturong (*Arctictis*), which are not far from the stem viverrines except in regard to certain specializations; (7) the small-toothed palm civet (*Arctogalidia*), related to the paradoxurine group; (8) the African palm civet (*Nandinia*), which we believe is much more closely related to the paradoxurine branch than is admitted by Pocock; (9) the predaceous fanaloka, or fossane (*Fossa*) of Madagascar, which should not be confused with the cat-like true fossa (*Cryptoprocta*), since its dentition and skull tend to connect it rather with the common stem of the paradoxurid and hemigalid sections; (10) the Hemigalida, as defined by us, comprising the irregularly colored *Hemigale* and *Diplogale*, the fantastic *Chrotogale*, all of eastern Asia, the insectivorous *Proteles* of Madagascar with its enfeebled jaws and teeth, together with the otter-like *Cynogale* of Borneo; (11) the galidictine stock of Madagascar, which seems to be a mongoose-like derivative of the stem of *Fossa;* (12) the

[1] Gregory, William K., 1936, "On the Phylogenetic Relationships of the Giant Panda (*Ailuropoda*) to Other Arctoid Carnivora." *Amer. Mus. Novitates*, No. 878.

true mongoose family (Herpestidae), which is composed of about a dozen genera of snake-fighting and insect-eating forms found mostly in Africa, but with the typical genus *Herpestes* in Asia.

The classic investigations of Mivart (1882) on the anatomy, classification and distribution of the aeluroid Carnivora have been constantly before us as we have studied our material; nor have we neglected to compare and check our findings against the results obtained by Pocock in his numerous important papers. These deal with a wide range of morphological data, including many external features (rhinarium, vibrissae, ears, color patterns of head and body, details of the digits and digital pads, etc.), together with the external anatomy of the perineal and circumanal glands and of the genitalia.

Mr. Pocock's studies have led him to place a high systematic value on the variations in the characters and position of the scent glands throughout the viverroid Carnivora. He is of the opinion that these organs are of such high physiologic importance that it is difficult to conceive a rapid transition from their presence in the typical Viverrinae to their absence in the linsangs. He is thus led to the use of them as his prime criterion in dividing the family Viverridae into numerous subfamilies. He regards the skull and dental characters as being in general more variable and less important as indicating degrees of relationship than the differences in the scent glands, as above noted. He makes an exception, however, in the case of the auditory region and adjacent parts in the skull of the African genus *Nandinia*, which he has removed altogether from the vicinity of the palm civets, where Mivart placed it, giving it the rank of a separate family, Nandiniidae, coördinate with the Viverridae, Mungotidae (Herpestidae), Cryptoproctidae, Hyaenidae, and Felidae.

We regret to state that, although we find Mr. Pocock's factual observations of the greatest value, we have been compelled by our own experience to conclude that there are several well marked instances (notably among the paradoxures and arctogalidians as recorded by Pocock) where the perineal glands give evidence of transitional conditions in otherwise closely related genera; we also find that the peculiarities of the auditory region of *Nandinia* are significant only of generic rank, not of family divergence from the paradoxurine stock. We have moreover been led to certain general conclusions, more fully stated at the end of this paper, to the effect that the doctrine of "irreversibility of evolution" has been pushed to extremes by many mammalogists and palaeontologists, who have thereby been prevented from realizing that in the course of evolution there have been retreats as well as advances,

and that the specializations of earlier ages have often been wiped out or covered over by the specializations of later ages.

With regard to methods we have constantly and repeatedly compared the skull and dentition of each genus with all the other genera, related and unrelated, of the aeluroid series, seeking by the method of contrast and comparison to discover what the differences are, which differences are more constant and which more variable, which are relatively new "habitus" features and which are older "heritage" characters. In the course of this analysis we have made many hundreds of comparative measurements and indices, but in the present paper we are citing only a very small part of our metrical data.

For the material studied in the present paper we have to thank primarily Dr. Harold E. Anthony, Curator of the Department of Mammals in the American Museum of Natural History, who generously placed at our disposal the large collection of aeluroid Carnivora under his care. A good part of this material was collected by Herbert Lang during the Lang-Chapin Expedition to the Belgian Congo in 1909–1915. Recently important viverrid material from Madagascar has been collected by Mr. Richard Archbold and Mr. G. H. H. Tate of the Rand-Archbold Expeditions. Other curators who have kindly allowed us to study valuable material are Dr. Wilfrid S. Osgood, Chief Curator, Department of Zoology, Field Museum of Natural History, Chicago, Dean Houghton Holliday of the College of Dental and Oral Surgery, Columbia University, and our colleague Mr. Henry C. Raven, who collected examples of *Nandinia* and *Xenogale* in the French Cameroons, West Africa. We are obliged to our colleague Dr. Edwin H. Colbert of the Department of Vertebrate Palaeontology of the American Museum for access to the creodonts and other fossil Carnivora described by Cope, Matthew and others.

We may begin by presenting a brief major classification of the Carnivora herein adopted and may then proceed to consider the families and other major divisions of the aeluroid group.

Since we are concerned only with generic differences, we have refrained from giving specific identification in many cases throughout this paper.

Major Classification of the Carnivora

Class Mammalia
  Subclass Theria
    Infraclass Monodelphia (Placentalia)
      Order Carnivora

[Suborder Creodonta Cope (*partim*)]
Infraorder Procreodi Matthew [Superfamily Arctocyonoidea Simpson]
Family Arctocyonidae (including Oxyclaenidae, Triisodontidae)
Infraorder Acreodi Matthew [Superfamily Mesonychoidea Simpson]
Family Mesonychidae
Infraorder Pseudocreodi Matthew [Superfamily Oxyaenoidea Simpson]
Family Oxyaenidae
Family Hyaenodontidae
Infraorder Amphicreodi [1]
Family Palaeonictidae
[Suborder Fissipedia auct. (Carnassidentia Wortman)]
Infraorder Eucreodi (Matthew)
Superfamily Feloidea Simpson (Aeluroidea Flower)
Family Viverridae (including Miacidae)
Family Herpestidae
Family Hyaenidae
Family Felidae
Superfamily Musteloidea *nobis*
Family Mustelidae
Superfamily Canoidea Simpson (*partim*, Arctoidea Flower *partim*)
Family Canidae
Family Procyonidae
Family Ursidae
[Suborder Pinnipedia]
Superfamily Otarioidea
Family Otariidae
Family Odobaenidae
Superfamily Phocoidea
Family Phocidae

The first noteworthy feature of this classification is the transference of Matthew's Eucreodi from the Creodonta to the Fissipedia, a step originally taken by Wortman, who proposed the name Carnassidentia to replace Fissipedia in the larger sense. Although Matthew recognized that the Procreodi were widely distinct from the Pseudocreodi and that

[1] New infraorder.

they were ancestral to all the modern families, his failure to remove them from the Creodonta to the Fissipedia was largely due to his conservative adherence to the "horizontal system" of classification as practised by Cope, in opposition to the "phylogenetic system," which he felt was being pushed to extremes by Osborn.

Our reasons for proposing a new Infraorder Amphicreodi are given below (p. 316).

Our removal of the Mustelidae from the Arctoidea and the insertion of this group as a superfamily lying between the Feloidea and the Canoidea (*s.s.*) is occasioned by evidence that the immediate ancestors of *Mustela*, in the lower Oligocene of North America, possessed a skull and dentition that were, on the whole, close to that of *Herpestes*.

The Origin and Rise of the Aeluroid Carnivora

In order to view the viverroid carnivores in their proper geologic perspective, let us look briefly at the palaeontologic history of their collateral ancestors among the creodonts. The late W. D. Matthew, developing the work of Cope and others on the mammals of the Paleocene, concluded (1909) that the oxyclaenid creodonts were the most central and primitive group of the Carnivora and that they were not far removed in structure from the common ancestors of the ungulate orders. From Matthew's studies (posthumously published 1937) it appeared to him probable that the skeletal construction of one of these oxyclaenids named *Loxolophus hyattianus* Cope "presents the closest known actual approach to the common primitive type from which the placental mammals, or most of them, are derived." The postcranial skeleton of this animal, which is about as large as a genet or a small cat, is of the primitive semi-arboreal type, with spreading, semi-digitigrade hands and feet, divergent, semi-opposable and unreduced pollex and hallux, well developed, probably non-retractile, claws and unrestricted power of pronation and supination of the forearm.

The skull of *Loxolophus*, which is about as long as that of a genet, has the braincase less expanded transversely, the postorbital constriction much longer, the muzzle longer. The basicranial region is much shorter anteroposteriorly and wider transversely. Although the basicranial region is imperfectly preserved in this skull, it is known from related forms that the auditory bullae, if present, were membranous (based on *Deltatherium*, an allied genus), leaving the auditory prominence of the periotic exposed in the fossil, whereas in the Viverridae it lies deep within the bony bulla. The paroccipital process, unlike that of modern viverrids, is not wrapped around the posterior surface of the

bulla. The cranial foramina, so far as known, are all of generalized placental type.

While *Loxolophus* is slightly specialized in the separation of the first and second upper and lower premolars from each other and from their neighbors, the outstanding feature of its dentition, next to its general

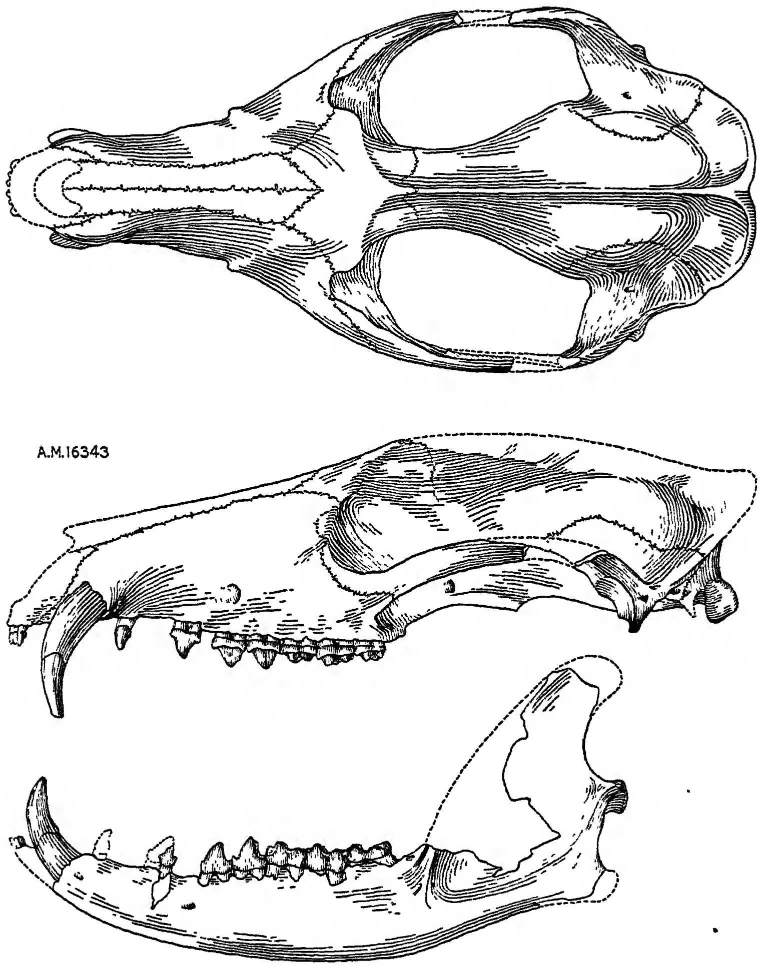


FIG. 2. *Loxolophus hyattianus*, skull and lower jaw. After Matthew. × 10/11.

carnivorous-omnivorous habitus, is the sharp difference between the last upper and lower premolars and the first molars. This, as G. G. Simpson has pointed out, is a feature of the Mesozoic Pantotheria and probably a primitive placental character; so also are the facts that the first molar is much larger and more complex than the last premolars in

both upper and lower jaws, and that the third upper molars are still well developed although smaller transversely than the second. The upper molars retain the triangular ground-plan but with well defined posterior cingulum, which bears, at least on m^1, m^2, an incipient hypocone. The second upper molar is larger than the first, whereas in the viverrids the reverse is the case. The last upper premolar (p^4) is a simple bicuspid with one large outer, and one smaller inner, cusp. In the lower molars the trigonids are distinctly higher than the talonids, a primitive condition; the talonids are wider than the trigonids on m_1, m_2; m_3 is a narrow, somewhat elongate tooth.

The adult dental formula of *Loxolophus* ($I\frac{3}{3}\,C\frac{1}{1}\,P\frac{4}{4}\,M\frac{3}{3}) \times 2 = 44$) is likewise that which has long been recognized as the "primitive Placental formula." Unfortunately, the deciduous dentition of Paleocene creodonts is unknown but to judge from its condition in recent carnivores (cf. Leche, 1915) there is no reason to suppose that it was other than $DI\frac{3}{3}\,DC\frac{1}{1}\,DPM\frac{3}{3}$, the first functional premolars in young stages being probably early erupted members of the permanent set.

Among the creodonts there is a wide adaptive variation in the dentition from the central type described above, leading at one extreme to the flattened molars of *Anacodon* and at the other to the shearing blades of the oxyaenids and hyaenodonts. It has been shown, especially by Matthew, that the extreme carnassial specialization of different families of creodonts affects chiefly the second upper molar in the Hyaenodontidae, the first upper molar in the Oxyaenidae, and the last upper premolar (p^4) in the Miacidae. Matthew regarded the Eucreodi as wholly independent of the Pseudocreodi, but in view of numerous resemblances between these groups in the skull and postcranial skeleton we should not be surprised if the Eucreodi had arisen from primitive Pseudocreodi simply by a forward shift of the maximum growth gradients from $\frac{m^1}{m_2}$ to $\frac{p^4}{m_1}$. The European and American Eocene genus *Palaeonictis* and its relative *Ambloctonus* have been classed by Matthew (1915, pp. 5, 57) with the Pseudocreodi because their "carnassial" teeth were m^1, m_2, as in the Oxyaenidae. But they also had good shearing blades on p^4 and on m_1, m_2. To transform the dentition of a primitive palaeonictine, such as *Dipsalodon mathewi* Jepson (1930, p. 524), into that of the stem miacid would require mainly that there should be a progressive diminution of m^1, m^2, especially in the transverse direction, of m_2, m_3 in the anteroposterior direction, in a manner precisely similar to that which is known to have occurred in various lines of eucreodine Carnivora in which the carnassial function became predominant. In any case,

the subfamily Palaeonictinae Denison (1938, p. 174), although referred to the Pseudocreodi, is at most an outlying member of that group and it is certain that its dentition more closely approaches the eucreodine stage

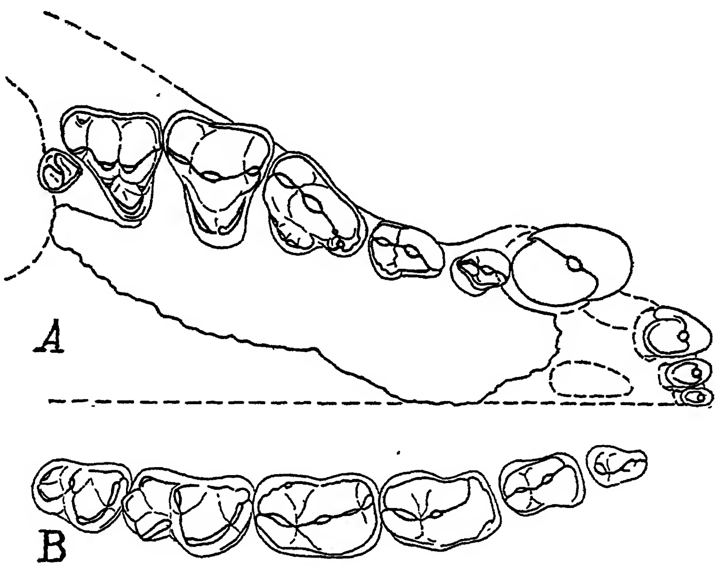


FIG. 3. *Palaeonictis occidentalis*, upper and lower cheek teeth. Occlusal view. After Denison. *A*. Natural size. *B*. Somewhat enlarged.

than does that of any known oxyclaenid, in all of which p^4 was much smaller than m^1 and had not yet acquired a large shearing metastyle blade. The limb and foot structures of the more primitive of the Pseudocreodi (e.g., *Thinocyon*), as recently described by Denison (1938), are very primitive and apparently suitable to be structurally ancestral to the conditions in the Eucreodi and later Fissipedia. In view of the structurally transitional characters of *Palaeonictis* and its allies between the stems of the Pseudocreodi and the Eucreodi, we herewith remove the Palaeonictinae from the Pseudocreodi and erect for them an intermediate group coordinate with the Pseudocreodi and Eucreodi, which may be named Amphicreodi and be defined by the fact that the carnassial blades on p^4, m_1 are larger than those on m^1, m_2.

Whether the Eucreodi and with them the typical fissipedes were derived, as Matthew assumed, from the very primitive Procreodi, or, as we suspect, from relatives of the amphicreodine subfamily Palaeonictinae, it appears highly probable to us that those forms of later Carnivora, with little or no carnassial blade on p^4 and blunt conical cusps on the molars, were in all cases derived from the viverrine or subcarnassial

stage rather than the reverse, and in the present paper we shall set forth the reasons that have led us to this conclusion.

Whether the carnassial development takes place on p^4 or m^1 or m^2, the metastyle blade always becomes emphasized and so does the

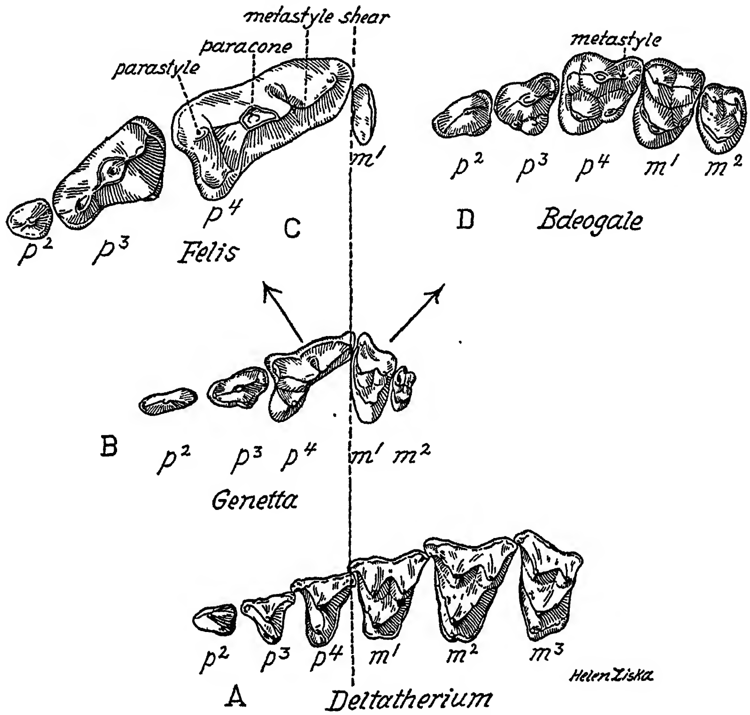


FIG. 4. Morphology of the upper carnassial teeth.

A. Primitive Eocene stage (*Deltatherium*), with incipient metastyle shear on p^4 and unreduced molars.

B. *Genetta*, showing emphasis of metastyle blade, loss of m^3 and reduction of m^2.

C. *Felis*, showing extreme emphasis of metastyle shear, loss of m^3, m^2, and severe reduction of m^1.

D. *Bdeogale*, procyonoid derivative of a *Genetta*-like stage, showing reduction of metastyle of p^4 and enlargement of inner cusps on p^4, m^1, m^2.

paraconid-protoconid shear of its antagonist of the lower jaw. Meanwhile the paraconid-protoconid shear of the lower carnassial (Fig. 5) swings from a more transverse to a more anteroposterior position. The internal ("protocone") cusp of the upper carnassial tooth is displaced forward; its inner side acts, together with the talonid of its antagonist, the lower tooth in front of it, as a stop or holdfast for the food in front of the enlarging paraconid of the lower shear tooth. Thus in the typical Miacidae of the lower Eocene, which are the immediate forerunners of

the viverroid Carnivora, the internal cusp of the carnassial p^4 is small, the tooth is narrow transversely, its metastyle blade elongate posteriorly.

In *Didymictis* and *Viverravus* of the Miacidae, which represent a major stage of advance beyond the oxyclaenids, the carnassials (p^4 and

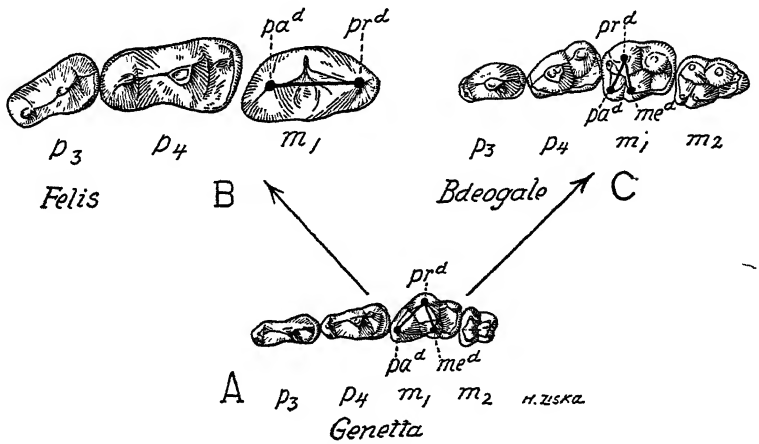


FIG. 5. Morphology of the lower carnassial teeth.

A. Genetta, showing incipient enlargement of paraconid and protoconid blade.

B. Felis, showing extreme enlargement and anteroposterior direction of paraconid-protoconid blade, with complete loss of talonid; loss of m_2.

C. Bdeogale, showing tubercular development of main cusps, secondary approach of paraconid and metaconid, increase of talonid, enlargement of m_2.

m_1) are larger than the molars and the third molar is small or absent in both upper and lower jaws. In the Miacidae no less than in their predecessors the Oxyclaenidae (Arctocyonidae) we find a wide range in the size of the carnassial, which varies inversely from that of the molars, especially when the latter are becoming flat-topped and massive as in *Palaearctonyx.*

The postcranial skeleton of the Miacidae, as described by W. D. Matthew in his "Memoir on the Carnivora and Insectivora of the Bridger Basin," varies around a central type with grasping hands and feet, which Matthew interprets as a primitive arboreal or semi-arboreal type. Of this family *Viverravus* and *Didymictis* constitute the subfamily Viverravinae, which may be ancestral, or at least closely related, to the ancestors of the Viverridae and their derivatives, the hyaenas and cats, while the Miacinae tend as a rule to develop the hypocone cingulum of the upper molars and are close to the most primitive members of the Canidae, which in turn gave rise to the Ursidae and Procyonidae. For reasons shown below, we regard the entire family "Miacidae" merely as a subfamily of the Viverridae.

Classification of the Sections and Subfamilies of the Viverridae and Allied Families

Superfamily Feloidea Simpson (Aeluroidea Flower)
- Family Viverridae
  - Section Viverrida *nobis*
    - Subfamily Miacinae (Miacidae Matthew *partim*)
    - Subfamily Viverrinae: *Viverravus, Didymictis, Viverra, Viverricula, Civetictis, Genetta, Osbornictis, Poiana*
    - Subfamily Prionodontinae [Pocock]:[1] *Prionodon, Pardictis*
  - Section Paradoxurida *nobis*
    - Subfamily Nandiniinae[2]
    - Subfamily Arctogalidiinae [Pocock]
    - Subfamily Paradoxurinae [Pocock]
  - Section Hemigalida *nobis*
    - Subfamily Fossinae [Pocock]
    - Subfamily Hemigalinae [Pocock]
    - Subfamily Cynogalinae [Pocock]
    - Subfamily Euplerinae [Pocock]
  - Section Galidictida *nobis*
    - Subfamily Galidictinae [Pocock]
- Family Herpestidae [Pocock]
- Family Hyaenidae
  - Section Protelida *nobis*
    - Subfamily Protelinae[2]
  - Section Hyaenida *nobis*
    - Subfamily Hyaeninae
- Family Felidae
  - Section Cryptoproctida *nobis*
    - Subfamily Cryptoproctinae[2]
  - Section Felida *nobis*
    - Subfamily Pantherinae Pocock
    - Subfamily Felinae Pocock
    - Subfamily Acinonychinae Pocock
  - Section Macherodontida *nobis*
    - Subfamily Macherodontinae auct.

[1] In this and the following table the name of an author in square brackets following the name of a subfamily is intended to mean only that we are following that author's usage as to the content and limits of that subfamily.

[2] Treated as a family by Pocock.

The most conspicuous feature of the foregoing classification of the aeluroid families is the insertion of a grade called *section*, just below the family. The section may include one or more of the subfamilies recognized by Pocock. Since these sections are nearer in principle to the family than to the subfamily, they are given the arbitrary termination *ida*, rather than *ini*, which has been used in a somewhat similar way by Childs Frick (1937) in grouping the numerous subfamilies of the Tertiary horned ruminants. Our sections rest in each case primarily upon the evidence, to be cited below, tending to suggest derivation of all the included subfamilies from a common stock.

The Ancestry of the Civets and Genets

Family Viverridae

Eucreodine carnivores, in which the auditory bulla is composite (composed of ecto- and entotympanic moieties), although not always fully ossified (?*Didymictis*, *Nandinia*). Third upper molar reduced or, usually, absent. Internal cingula of upper molars slight or absent. Carnassials, p^4, m_1, primitively sectorial, becoming variously modified (feline, procyonine, myrmecophilic, etc.). Neck typically long, back arched, tail long; feet typically subcursorial, subdigitigrade, with more or less retractile claws. Gait slinking to cursorial, secondarily scansorial, semifossorial or incipiently natatorial. Scent glands usually present in the perineal region (Pocock).

Section *Viverrida*

The central stock of the Viverridae: skull and dentition carnivorous, cheek teeth viverrine; auditory bulla composite, with both ecto- and entotympanic ossified (except possibly in the Miacinae), the entotympanic if present becoming elongated anteroposteriorly with the paroccipital process wrapped around it; cranium narrow and long; bony muzzle narrow and low (except in Poiana); postorbital constriction pronounced to moderate. Feet pentadactylate, typically cursorial, usually with more or less retractile claws (except in the Miacinae). Perineal scent glands usually present (except in the Prionodontinae).

In 1909 Matthew arranged the eucreodine genera of the Eocene of North America into two subfamilies, which he named respectively Viverravinae (*Viverravus, Didymictis*) and Miacinae. The American Paleocene *Didymictis* is a very central type of the Eucreodi. On the one hand, it retained numerous primitive creodont skeletal characters, including the unexpanded braincase and the separate elements of the scapho-lunar-centrale bone of later carnivores; on the other hand, it had

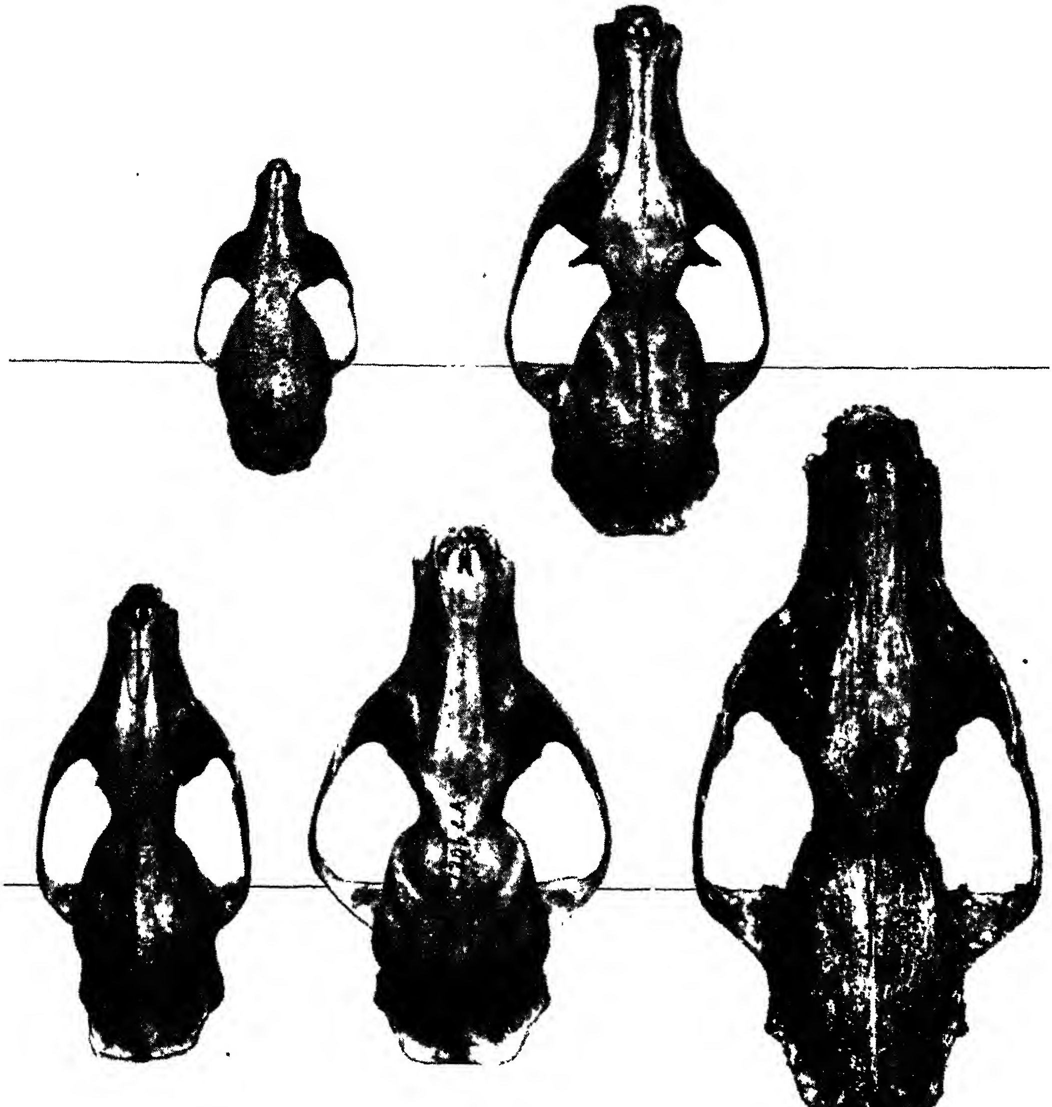
already acquired a completely eucreodine stage in having fully developed "carnassials" on p^4 and m_1. The dentition of *Ictidopappus* Simpson (1937, pp. 213, 214), *Didymictis* and *Viverravus* is fundamentally the same as in *Viverra,* although somewhat more primitive in details. On the evidence of the dentition we refer to the Viverridae at least the "miacid" genera mentioned above, although their auditory region had possibly not yet acquired a conspicuous feature that is diagnostic of modern viverrids, namely, the presence of an ossified compound auditory bulla composed of ecto- and entotympanic moieties.

Apparently the bony bulla was first formed exclusively of the ecto-tympanic bone, or tympanic proper. This is the condition retained in the majority of placental mammals of other orders as well as in all the so-called arctoid families. In whatever way the entotympanic division of the bony bulla originated, it at least formed the outer shell of a membranous bubble that was connected by a foramen with the epi-tympanic sinus of the tubo-tympanal or Eustachian canal. By lower Oligocene times the "Stenoplesictoïdés" of Teilhard de Chardin contained certain forms in which the characteristic bipartite viverrine bulla is almost or completely realized.

In 1914–1915, Teilhard de Chardin in the course of a most vigorous and comprehensive analysis of the available evidence divided the numerous species and genera of the miacid Carnivora of the "Phosphorites du Quercy" into four main groups, which, however, are not of equal grade (1915, p. 5): (1) the "Miacoïdés," including *Miacis exilis, Viverravus angustidens* and *Cynodon miacinus,* very primitive forms related to the older Miacidae of North America; (2) the "Cynodictoïdés," including a wide range of species and genera, of which the more central forms may have led to some of the amphicyons or giant dogs; still other Cynodictoïdés seem to have led to (3) the "Cynodontoïdés," including, among many others, possible ancestors of certain of the Mustelidae and Procyonidae; (4) a second main division of the Cynodictoïdés was the "Sténoplésictoïdés," which was in turn subdivided into a "Branche Viverrienne" and a "Branche Mustélienne," the former including *Palaeoprionodon* and related genera in which the skull had attained all the diagnostic characters of the modern Viverridae.

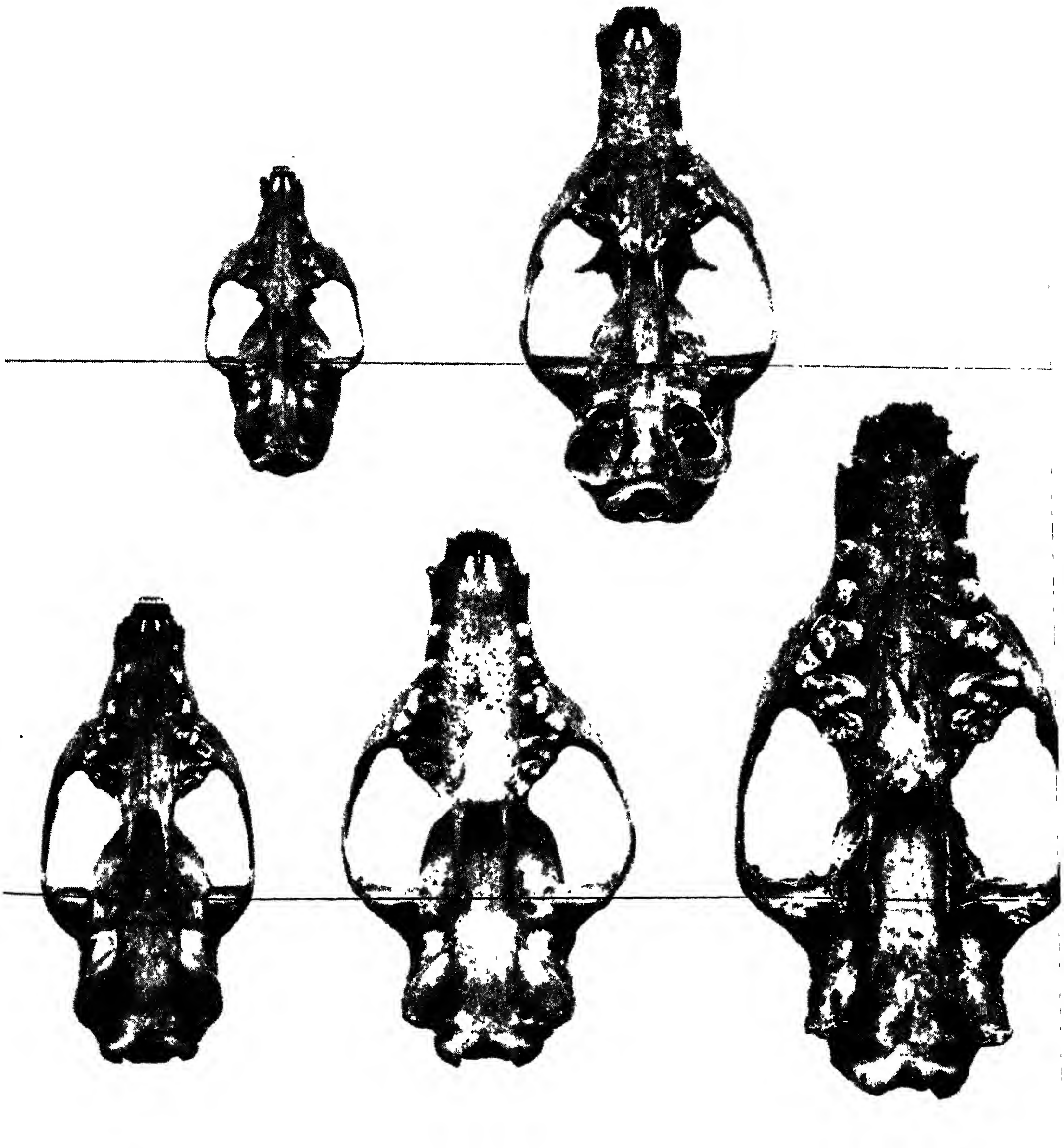
Of the "Miacoïdés," *Miacis exilis* (Teilhard, 1915, Pl. I, Figs. 1, 5) is a very small form retaining a reduced third lower molar and with an insectivore-like, pricking trigonid on m_1; the talonids of both m_1 and m_2 are much less expanded than in modern Viverrinae and Hemigalinae. In *Viverravus angustidens* (*op. cit.*, Pl. I, Figs. 6, 7, 9–12) the skull is a little shorter than those of the modern *Viverra zibetha* of India figured

PLATE I



SKULLS OF THE VIVERRINAE. Top view. × ⅔. Upper row, left to right: *Poiana, Osbornictis.* Lower row, left to right: *Genetta, Viverra, Civetictis.*

PLATE II



SKULLS OF THE VIVERRINAE. Underside view. × ⅔. Upper row, left to right: *Poiana*, *Osbornictis*. Lower row, left to right: *Genetta*, *Viverra*, *Civettictis*.

by Pocock (1933, p. 439) but it is much more primitive in many characters: (1) it is a much more robust, broader skull, with heavier crests; (2) the dentition is more massive; (3) the first upper molar is more dog-like, less compressed anteroposteriorly; (4) the large secant m_1 has a much smaller talonid than in the modern *Viverra*; (5) m_2 is very small (from this and other evidence we are persuaded that the expanded tuberculated talonid of m_1, as well as the flat-topped wrinkled surface of m_2, in the modern *Civetictis* is quite secondary); (6) the base of the cranium is relatively much shorter than in *Viverra*; (7) the entotympanic or posterior division of the bulla, if ossified, could not have been greatly inflated posteriorly as it is in modern Viverrinae, because the well developed paroccipital process is directed backward much as it is in the recent *Nandinia* and therefore could not have been closely appressed to an inflated entotympanic. In short, as regards the general construction of the skull and dentition, the Miacoids of the Phosphorites, as well noted by Teilhard de Chardin, are on the whole distinctly less advanced in specialization than are the modern Viverrinae.

In the typical cynodictoids (Teilhard, 1915, Pl. III, Fig. 13) the most outstanding features of the skull and dentition are as follows: (1) the skull is relatively short and broad with a short basicranial region and probably unexpanded entotympanic; (2) the palate is short and wide; (3) the upper carnassial p^4 has the inner cusp small and located at the anterior end of the crown; the blade is fairly long (this is the type of carnassial which has been accentuated in the modern *Prionodon* and still more in *Cryptoprocta*); (4) the first upper molar has a large internal cingulum, much larger even than in the modern *Civetictis*; (5) the second upper molar is small but likewise bears a prominent internal cingulum.

The various species of *Cynodictis* and allied genera indicate a drift from the primitive Viverravinae to the earliest Canidae. The postcranial skeleton of the American Oligocene *Cynodictis gregarius*, according to Matthew (1899, p. 123), more closely resembles that of the existing *Viverra* than that of the true dogs. Thus it has the semiretractile claws and short hands and feet of the viverrids rather than the compressed hands and feet and nonretractile claws of the cursorial dogs, the latter characters being a later specialization.

In general the cynodictoids are transitional in range of characters between the civet-hyaena-cat group on the one hand and the dog group on the other. As already noted, the upper molars have a more dog-like development of the internal cingulum than do those of the civet-cat series, while the auditory region of the skull lacks the special viverroid characteristics that are conspicuous in the stenoplesictine group so far as known (see below, p. 324).

The skull of *Palaeoprionodon Lamandini* (Teilhard de Chardin, 1915, Pl. IX (XX), Figs. 10, 14), a representative of the stenoplesictoids, is identical in general plan, though not in details, with that of the existing viverrid genus *Fossa*. In the auditory region the ectotympanic, or true tympanic, bone was ossified, while the entotympanic was either membranous or very thin, so that, as observed by Teilhard, the cochlear portion of the periotic is exposed in the dried skull exactly as it is in the recent viverroid *Nandinia*. The entotympanic was evidently inflated, especially at the postero-external border, just as it is in modern viverrids. The frontal pole of the cerebrum evidently extended much farther forward than in the ancestral oxyclaenid creodonts, so that the postorbital constriction was not tubular but, as seen from above, presented the appearance of a groove behind the postorbital process of either side, much as it does in the existing genera *Fossa* and *Genetta*. The brain, however, seems to have been relatively shorter and wider than in the genera named. The muzzle was shorter and less delicate than the corresponding part in *Fossa*.

The dentition of *Palaeoprionodon* (*op. cit.*, Pl. IX, Fig. 10) is marked by the elongation of the blade of the carnassial, the transverse narrowness of the latter, by the reduction of m^1 to a small transversely placed tooth and by the absence of m^2. In these and other respects the dentition of *Palaeoprionodon* closely approaches not only its contemporary *Proailurus* (*op. cit.*, Pl. IX, Figs. 12, 18) but also its modern structural descendants *Prionodon* and *Cryptoprocta*. In the lower jaw (*op. cit.*, Pl. IX, Figs. 7, 8, 8a) the first lower molar forms a prominent carnassial with reduced metaconid and short heel, while m_2 is much reduced and m_3 absent. Here again *Palaeoprionodon* is structurally related on the one hand to the supposed ancestors of the cats (*op. cit.*, Pl. IX, Figs. 19, 20), and on the other to the modern *Prionodon* and *Cryptoprocta*. Few palaeontologists with sufficient practice in assessing degrees of relationship as measured by degrees of identity in fundamental patterns would hesitate to endorse Teilhard de Chardin's tentative conclusion (p. 87) that the animal represented by the skull of "*Stenoplesictis* (*Cayluxi*)" was approaching the modern *Prionodon*.

In short, the placental carnivores of the Phosphorites (lower Oligocene) of Europe show a wide diversity of details in the pattern of the dentition and in the construction of the skull. In general, the cynodictoid division seems to be tending toward the dogs, the cynodontoids toward the weasels (mustelids), the stenoplesictoids toward the civets (viverrids), hyaenas and cats. But all were so closely related to each other in the development of the carnassials on p^4 and m_1 and in the

general construction of the skull that it would seem artificial to distribute them into different superfamilies; indeed, so far as we can infer from his text, such was not Teilhard de Chardin's intention in coining the French words "Cynodictoïdés," etc.

In general, each successive grade of organization among the placental Carnivora shows in the upper and lower cheek teeth a structural range that includes on the one hand forms with sharp-cutting V's on their molars, and on the other, forms in which the cutting edges are modified by (*a*) the development of a hypocone, (*b*) the transformation of the V's into rounded cones and (*c*) the upgrowth and secondary cuspidation of the hypoconids.

This general story is told first in the Paleocene and lower Eocene Oxyclaenidae, next in the Paleocene and lower and middle Eocene Miacidae, third, among their descendants of the lower Oligocene of France, fourth, among the modern civets and their allies. However, the earlier transformations did not attain such extremes as may be seen in the latest. As to geographic distribution, the Oxyclaenidae were common to Europe and western North America, as were also the Miacidae; the modern Viverridae, Herpestidae and their allies are found only in Europe, Asia, Africa and Madagascar. Although none of the Viverridae proper are found in North America, two of the derived families, namely, the Mustelidae and the Canidae, flourished in North America as well as in Europe. The origin and distribution of the Canidae, Procyonidae, Aeluridae, Ursidae, are discussed elsewhere (Gregory, 1936).

The more central genera of the modern Viverrinae, including *Genetta*, *Viverricula* and *Viverra*, have on the whole departed but little from the primitive viverrid stock of the lower Oligocene described above. Their outstanding characteristics are: (1) the high development of the perineal scent glands; (2) the marked anteroposterior elongation of the entotympanic chamber of the compound bulla; (3) the carnassial form of the cheek teeth. As to the last, it would ordinarily be assumed that the conditions in *Genetta* are all more primitive characters than those in the lower Oligocene viverrids, since the blade of its upper carnassial is less elongate and more outwardly directed than the corresponding parts in the extinct forms and the internal spur (protocone) is less anterior in position. To us, however, after observing apparent drifts away from a more carnassial dentition to various less carnassial, more tuberculate types of cheek teeth, in numerous lines of creodonts and fissipedes (*e.g.*, in the Oxyclaenidae, Canidae, Procyonidae, Mustelidae, etc.), it seems on the whole much more probable that the upper carnassial of the

primitive viverrid was more dog-like, that is, more *Cynodictis*-like, in general character than in any modern viverrid and that the structural contrast between p^4 and m^1 was much greater in the earlier forms than it is in such (to us) relatively specialized types as *Cynogale*, *Arctogalidia*, *Suricata*.

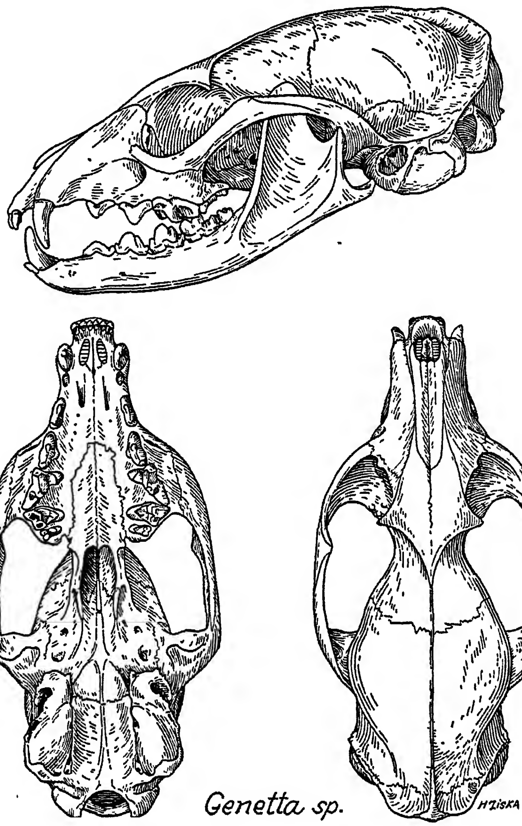


FIG. 6. *Genetta*, three views of skull. × ¾.

Among the Viverrinae we seem to have a morphological series leading from the Phosphorites *Viverra simplicidens* to the modern *Civetictis*, in which there has been a secondary increase in size of m^1 and even of m^2, together with a tendency toward flattening and roughening of the surface of the crown. The dentition of certain specimens referred to *Civetictis* is indeed almost dog-like in general appearance but a closer inspection reveals the lack of such diagnostic canine features as the markedly asymmetrical development of the internal cingulum shelf on m^1, m^2, so that *Civetictis* has a dog-like habitus but a *Viverra* heritage. In *Viverricula*, on the other hand, the upper molars and carnassials are much more delicate, more like those of *Palaeoprionodon*.

In *Osbornictis piscivorous* Allen (1924) the general appearance of the skull is somewhat like that of *Genetta* but there are many differences in detail, especially the following: the braincase is relatively wider, the basi-occipital segment relatively shorter, the zygomata more spreading,

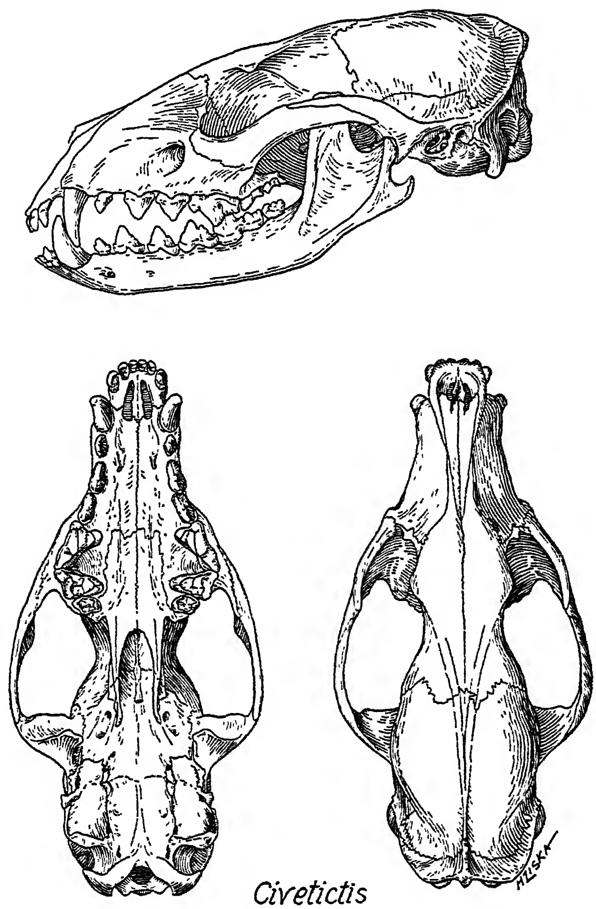


FIG. 7. *Civetictis*, three views of skull. × ½.

especially posteriorly, the compound bullae wider posteriorly, the muzzle and palate longer, the postorbital processes much more prominent, the forehead wider, the sagittal crest less prolonged posteriorly, the upper carnassial p^4 relatively larger and m^1 relatively smaller. The lower carnassial is normal and m_2 small.

In the postcranial skeleton Allen notes numerous minor differences from *Genetta victoriae*, presumably all specializations in connection with the fish-catching habits of the animal. The nearly uniform reddish

pelage is markedly different from the prominent spots and gray background of *Genetta*.

In spite of its fish-eating habits the skull of *Osbornictis* shows but few special resemblances to that of *Cynogale*, the otter civet: for example,

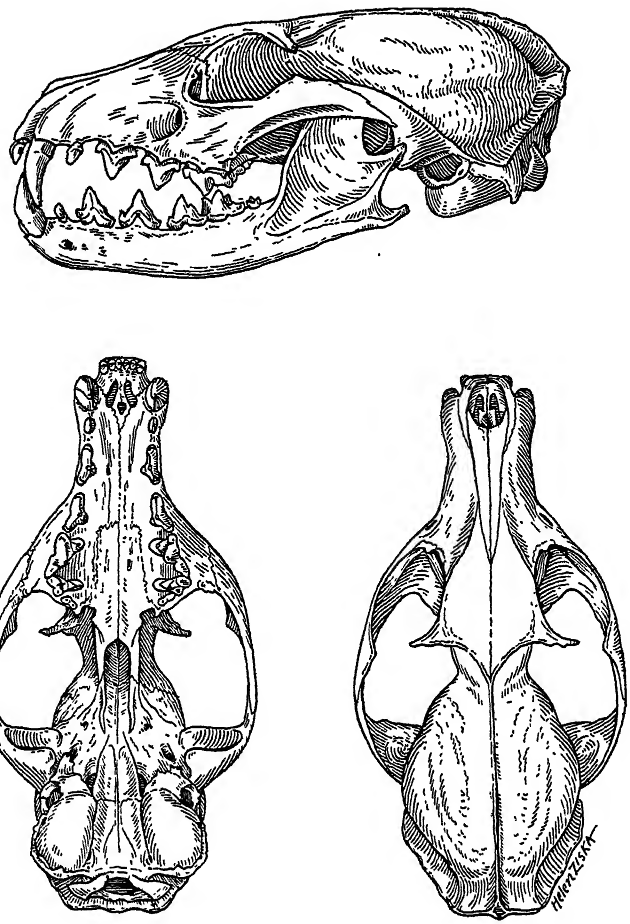


FIG. 8. *Osbornictis*, three views of skull. × ⅔.

the somewhat elongated muzzle, which is pinched dorsally between the prominent fossae for the strong muscles of the snout; large orbits with prominent postorbital processes, relatively wide occipital plane, bullae expanded posterolaterally. Along with such resemblances go differences corresponding both to the different relationships of these two genera and to differences in food habits: here belong the sharply carnassial cheek teeth of *Osbornictis* in contrast with the blunt-cusped rounded molars of *Cynogale*, the primitive lower premolars of *Osbornictis* with the serrated

premolars of *Cynogale*; the more primitive forehead of *Osbornictis* as against the extremely narrow forehead and postorbital constriction of *Cynogale*.

In short, *Osbornictis* is merely a moderately specialized side branch of the Viverrinae with certain convergent resemblances to *Cynogale*, which belongs in the hemigalidean division of the family Viverridae.

TABLE 1

VIVERRINAE

Skull Proportions	Index [1]	*Genetta*	*Viverricula*	*Viverra zibetha*	*Civetictis*	*Osbornictis*	*Poiana*	Remarks
Total length skull		82	89	132	147†	100.5	60.2*	Very wide range
Bizygomatic width	I	50	48.3*	51.5	52.4	54	56.1†	† Slight widening
Max. cran. width	VI	54.7	50*	54.1	60.5	60.3	60.8†	† Slight widening
Forehead width	VIII	41.4	41.4	55†	43.5	37.7*	41.7	† Slight widening
Interlac. width	II	61.5†	56.7	56.1	56.7	47.7*	60	* Possible narrowing
Palatal width	X	73	63.4	68.7	71.8	62.7*	76†	Opposite tendencies
Lower dent. arch width	XIV	54.1	53.8	48.5	55.4†	48.1*	50	Opposite tendencies
Lower jaw spread	XII	49.1†	45.9	38.4*	42.6	42.9	48.8	Opposite tendencies
Bicondylar width	XVII							
Olfactory length	III	54.7	53.4	75.7	93.1†	81	52.5*	† Great increase
Orbital length	IV	34	25.9	29.3	25.0*	34.5	35†	* Decrease
Basicranial length	XI	47.2	44.8	45.9	48.7†	41.4*	46.8	*† Opposite tendencies
Upper carnassial (P^4) width	XVIII	70.5	70.3	67.2	77.9	53.7	55.7	Extreme in *Poiana*
Lower carnassial (m_1) length	XVIIIa	68.4	63.4	54.3	51.4	73.1	80	

The skulls of *Genetta*, *Viverricula* and *Viverra* look generalized and in the table above none has more than two "highest" or "lowest" marks, whereas the other three genera have five or six. These primitive Viverrinae as compared with other subfamilies have skulls of moderate length and narrowness together with a relatively primitive subcarnassial dentition.

The most peculiar-looking skull is that of *Poiana*, which is the smallest in absolute length and has high width indices, very short muzzle and large orbits. *Civetictis*, which is relatively large, has smaller orbits and large muzzle. *Osbornictis* has a relatively narrow forehead, narrow interlacrymal, palatal, and lower dental arch, and short basis cranii.

[1] These indices are taken from our tables of measurements, which will be published later.
* lowest, † highest.

The Ancestry of the Hyaenas and the Aardwolf (*Proteles*)

That hyaenas usually feed on the remains of the feast left by the lions is almost proverbial; but although they are normally lacking in

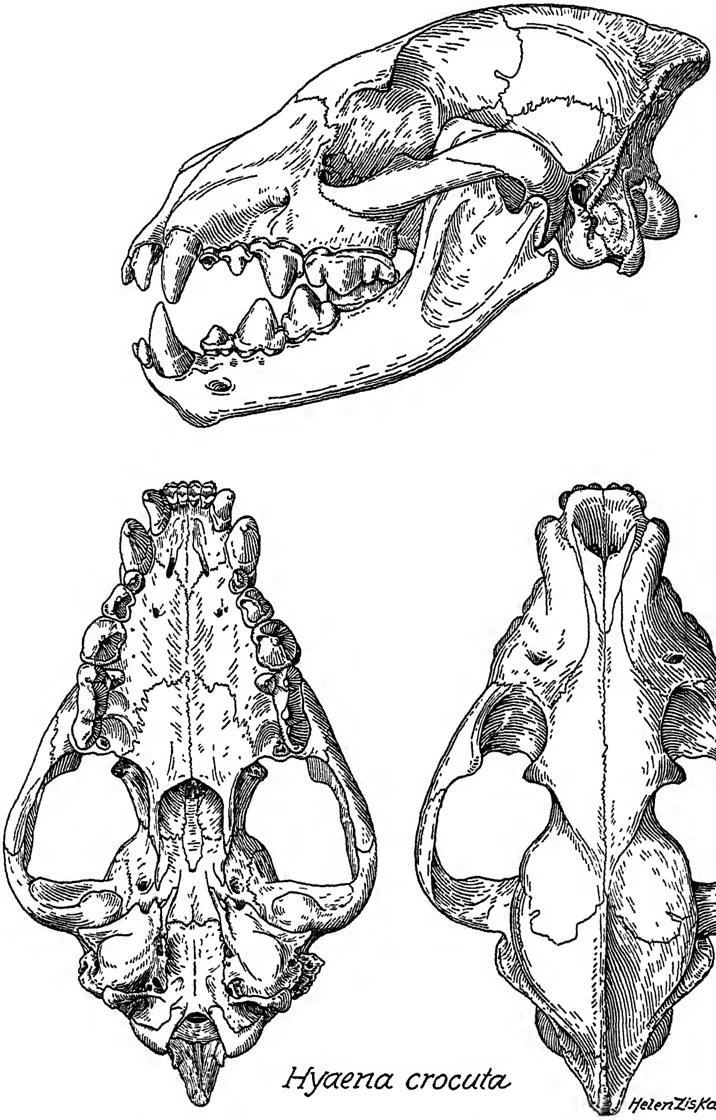


Fig. 9. *Hyaena crocuta*, three views of skull. × ⅓.

aggressiveness they have unsurpassed power in breaking bones with their jaws and shearing tough ligaments with their huge carnassial teeth. The massive upper carnassials have a large conical parastyle separated by a notch from the paracone, an unusually heavy metastyle shear and

a conical forwardly placed internal spur or protocone, all of which increases their breaking power. The huge temporal muscles secondarily restrict the brain cavity, giving it somewhat the appearance of certain small-brained creodonts. One outstanding specialization of the skull of the hyaenas is the marked reduction and posterior displacement of the entotympanic, as observed by Pocock (1916). The entotympanic has retreated with the advance of the ectotympanic, so that the conditions in the hyaenas present at first sight a deceptive resemblance to the bulla in the dogs and their allies, in which the ectotympanic alone forms the functional resonating chamber.

In the locomotor apparatus the hyaenas are far more specialized than typical viverroids, especially in their high forequarters and forelimbs, downwardly sloping back, relatively small hips and hind legs and much reduced tail.

The skulls of several species of the upper Tertiary genus *Palhyaena* (*Ictitherium*) furnish an unmistakable connection with the viverrine division of the Viverridae, as noted by Filhol and subsequent authors (*cf.* Boule and Piveteau, *Les Fossiles*, pp. 785). Thus the peculiar condition noted above of the tympanic region in the hyaenas is a good example of a transformation which has proceeded to such lengths that

TABLE 2

Proteles AND THE HYAENAS

Skull Proportions	Index	*Genetta*	*Proteles*	*Palhyaena*	*H. striata*	*H. crocuta*
Skull, total length		82	127	172	195	203
Bizygomatic width	I	50	63	65.7	76.9	?
Max. cran. width	VI	54.7	68.7	—	60.2	68.5
Forehead width	VIII	41.4	69.5	60.8	59.3	51.3
Interlac. width	II	61.5	60.4	55.8	56.0	57.9
Palatal width	X	73	74	76.7	83.3	90.6
Lower dent. arch width	XIV	54.1	67.3	55.7	58.4	70.8
Lower jaw spread	XII	49.1	57.1	40.9	44.4	53.4
Bicondylar width						
Olfactory length	III	54.7	98.5	—	99	88
Orbital length	IV	34	35.8	—	32.7	33.3
Basicran. length	XI	47.2	53	—	38.8	39.8
Upper carnassial width		70.5		—	63.0	56.3
Lower carnassial length		68.4		—	80.8	90.9

the descendant looks extremely unlike its remote ancestors. With regard to general external appearance, however, the hyaenas can now be recognized as specialized derivatives of such a typical viverrine as *Civetictis*.

The hyaenas, as compared with *Genetta*, have more than doubled the

skull length and made marked increases in nearly all transverse diameter ratios (bizygomatic, forehead, palatal, lower dental arch, lower jaw spread). The interlacrymal ratio is slightly diminished. Their olfactory length ratio has increased markedly, the basicranial ratio has decreased.

Palhyaena is intermediate in skull length between *Genetta* and *H. crocuta*. It approaches one or the other species of *Hyaena* in most of its measurements.

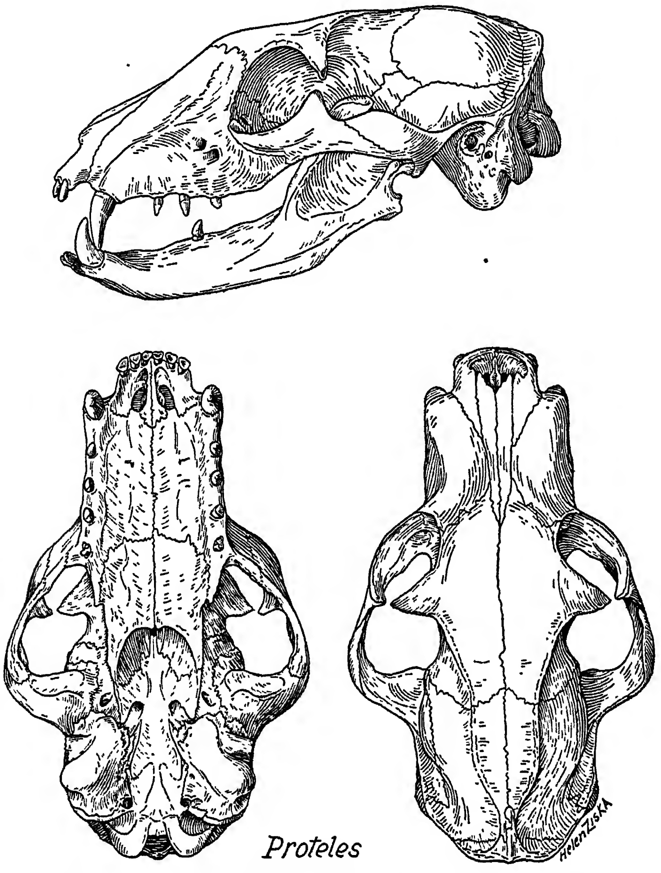


FIG. 10. *Proteles cristatus*, three views of skull. × ½.

Proteles, possibly because of its degenerate dentition and jaw muscles, does not agree closely in measurement with either the "primitives" or the hyaenas.

The aard wolf (*Proteles*), apart from its excessively degenerate dentition and very broad palate, much resembles a small hyaena which has taken to licking up termites as its major food. Its skull pre-

sents an interesting mixture of adaptive characters of different ages: (1) a recent termite-feeding habitus, manifested especially in the degenerate dentition and broad palate; (2) an older hyaena habitus, evidenced by the still powerful build of the skull, especially in the zygomatic arches, in the vertical deepening of the bullae and in the complete fusion of ecto- and entotympanic chambers; (3) remains of an early viverrine habitus, seen especially in the bipartite character of the bulla.

While the conclusion that *Proteles* is in effect a degenerate *Hyaena* seems best to accord with the evidence as we see it, we do not intend to pass over the observation of Mivart and Pocock that in a few characters *Proteles* recalls the herpestine group. Nevertheless, the following comparison of the auditory bullae and adjacent parts in *Genetta, Proteles* and *Herpestes* suggests to us that this important region affords evidence that *Proteles* is related to *Genetta* rather than to *Herpestes*.

	Genetta	*Proteles*	*Herpestes*
Maximum dorso-ventral depth of bulla	In middle of bulla	In middle of bulla, greatly emphasized	Distinctly posterior
Paroccipital process	Produced ventrally	Greatly produced ventrally	Not produced ventrally
Spout below bony auditory meatus	None	Pronounced	Incipient
Ventral slot in spout	None	None	Pronounced

In the absence of perineal scent glands *Proteles* agrees with both the herpestines and *Hyaena* but in possessing two circumanal glands it agrees with all three genera listed above.

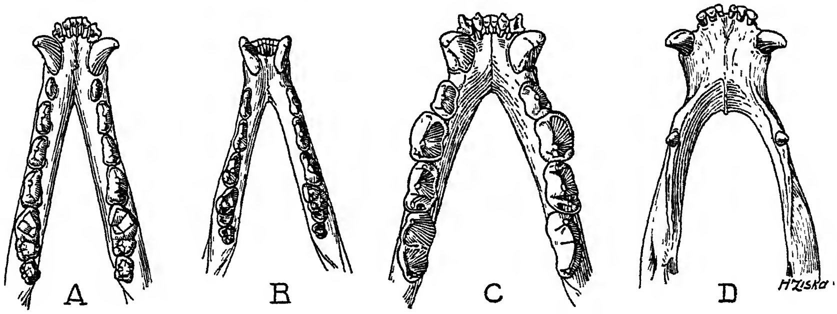


FIG. 11. Series of lower jaws. Scales various.
A. Viverra. B. Genetta. C. Hyaena. D. Proteles.

THE ANCESTRY OF THE LINSANG (*Prionodon*), THE FOSSA (*Cryptoprocta*) AND THE CATS (FELIDAE)

The several species of Asiatic viverrids called linsangs, constituting the genera *Prionodon* and *Pardictis*, were separated from the Viverrinae by Pocock (1915) on the ground that they lack the perineal scent glands which are so conspicuous in the latter subfamily. Moreover this genus approaches the cats in its glandless perineal area, in the posterior position of its external genitalia, in its hairy-soled feet, in its retractile claws and in the carnassial form of the dentition.

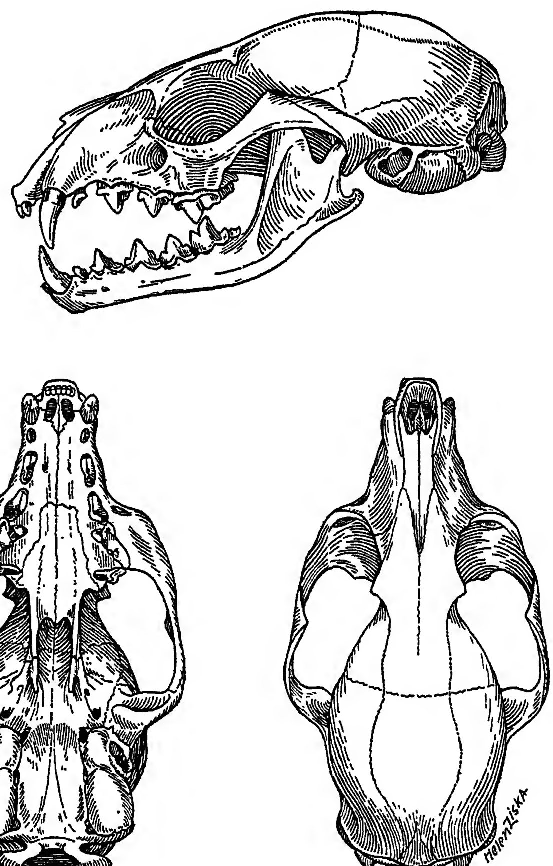


FIG. 12. *Poiana*, three views of skull. Natural size.

The African genus *Poiana* (Pls. I, II), formerly classified provisionally with the linsangs by Pocock, was afterward transferred by him to the Viverrinae, presumably because he had discovered that they have perineal scent glands.

Although both *Prionodon* and *Poiana* have an advanced carnassial type of dentition, they differ in respect to the following features:

	Prionodon	*Poiana*
Skull as a whole	Long and narrow	Broad
Braincase	Very narrow	Rounded
Sagittal crest	Very high	None
Entotympanic	Produced far forward	Not produced far forward
Ectotympanic	Small	Much inflated
Paroccipital process	But little expanded	Greatly expanded and wrapped around posterior side of bulla
External auditory meatus	Quite small, constricted	Very large and oval
Zygomata	Robust	Slender
P^1	Progressive	Retrogressive
M^1 anteroposterior diameter	Less compressed	More compressed

Part of these differences, however, may be due to the small size of *Poiana* and the large size of *Prionodon.*

The external appearance of the linsangs (*Prionodon*) as figured by Lydekker (New Natural History, I, Sect. II, p. 456) is not unlike that of the genets, except in details. The digitigrade tarsus of *Prionodon* as figured by Mivart (1882) and by Pocock, although remarkably elongated is fundamentally genet-like; nor does the manus seem sufficiently different from the genet type to indicate subfamily distinction. The claws are said to be nearly as retractile as in the cats.

The dentition is essentially like that of *Viverricula* but p^4 has a larger shear and m^2 is lacking (Mivart, 1882, p. 159). The resemblance of the extinct genus *Palaeoprionodon* of the upper Oligocene of Europe to the genus under consideration is implied in the name. Both the fossil and the recent genera exhibit a somewhat advanced stage of carnassial specialization (with long metastyle blade on p^4, m^1 much smaller than p^4, m^2 minute or absent, m_1 with talonid reduced or vestigial). The skull of *Palaeoprionodon* (Teilhard de Chardin, 1915, Pl. IX, Figs. 10, 14) may well represent the common ancestral type of all the diverse subfamilies of the Viverridae.

The bulla of *Prionodon* is of typical viverrine type, both the ecto- and entotympanic being well developed, but it does not seem to offer any conspicuous subfamily characters that would tend to distinguish it from members of the Viverrinae. Both the base of the cranium and the sagittal crest are elongated as in the Viverrinae; the paroccipital process does not form a thin close-fitting cap on the entotympanic but, at least in old specimens, bears a backwardly projecting ridge or prominence. In general the *Prionodon* skull closely resembles that of *Viverricula* (except in details), as do also the hands and feet.

In *Palaeoprionodon* the supposed absence of the entotympanic wall of the bulla, as figured by Teilhard de Chardin (1915, Pl. IX, Fig. 10), seems to be contradicted by the curvature of the paroccipital process as seen in the broken edge. In other words, the bulla of *Palaeoprionodon* was a composite, the ectotympanic forming the anterior moiety, the entotympanic (whether or not fully ossified) covering the posterior part.

The skull form of *Palaeoprionodon* is largely retained in the existing *Fossa*, while the dentition is nearly duplicated in the recent *Prionodon*, which is the type of Pocock's Prionodontinae.

The foregoing evidence seems to suggest that:

(1) *Prionodon* is a survivor of some lower Oligocene genus that was closely related to, or identical with, *Palaeoprionodon*, as noted by Teilhard de Chardin (1915, pp. 81, 87, and Pl. IX, Figs. 10, 14).

(2) Although the dentition of *Prionodon* is not far from that of the genets and other Viverrinae, it is apparently more advanced in the direction of the Cryptoproctinae.

(3) *Prionodon* and *Palaeoprionodon* may be regarded as representatives of a subfamily, Prionodontinae Pocock, characterized by reduction or absence of the perineal scent pouch (in the living forms), in combination with an essentially genet-like skull, a rather advanced carnassial dentition and elongated digitigrade feet with retractile claws.

(4) The subfamilies Viverrinae, Prionodontinae, as limited by Pocock, may be brought together as one of the main divisions (Viverrida) of the family, characterized by the possession of a carnassial dentition, an elongate cranial base and elongate compound bullae.

(5) The fossil genera *Palaeoprionodon* and *Proailurus* are close to each other, the former lying in or near the line of ancestry of *Prionodon*, the latter near to or in the lines, on the one hand, of *Cryptoprocta* and on the other, of the Felidae.

Family Felidae

Section *Cryptoproctida*

Cat-like derivatives of the stem Viverridae, with advanced carnassial dentition but retaining many primitive viverrid characters in the skull, backbone and limbs; perineal scent glands lacking; a circumanal pouch (into which glands open), present; external genitalia more "specialized" (?) than in the cats.

The fossa of Madagascar (*Cryptoprocta ferox*) has often been confused with the fanaloka (the genus *Fossa*), belonging to the subfamily Fossinae. It appears to be a little modified derivative of the European lower Oligocene *Proailurus*. It is a forest-living predator with a reputed

fondness for killing domestic fowl and an almost cat-like grade of advancement in the dentition, especially in the carnassials; of these, the lower carnassial (m_1) has wholly lost the metaconid and greatly reduced

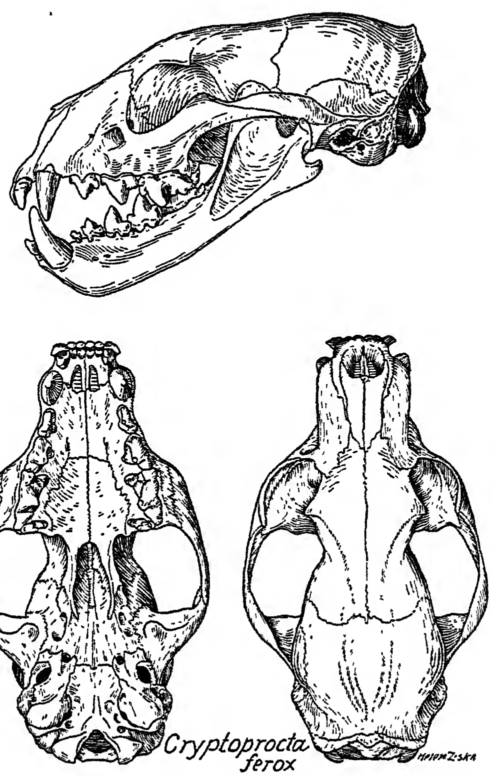


FIG. 13. *Cryptoprocta ferox*, three views of skull. × ½.

the talonid, while the upper likewise differs but little from that of typical felids.

On the other hand, *Cryptoprocta* retains a whole series of primitive viverroid characters in its backbone, hands, feet, and skull, which may be used if one so desires as evidence that this genus belongs in the Viverridae. With regard to its scent glands, Pocock shows that it lacks the perineal scent glands that are characteristic of typical Viverridae, while possessing a circumanal pouch into which glands open, resembling to some extent the condition in the hyaenas and mongooses. Pocock has also shown that the external genitalia of *Cryptoprocta* differ markedly from those of the cats; but after comparison of Pocock's figures of the external genitalia of *Cryptoprocta* with those of several preserved Felidae, it is at least conceivable that the conditions in the

latter may represent the combined results of degeneration and further specialization. The auditory region of the skull of *Cryptoprocta* has remained on a typical viverroid plane, since both elements of the compound bulla, the ecto- and the entotympanic, are subequally developed, whereas in the cats the entotympanic becomes greatly magnified.

Mivart (1882, p. 193) held that *Cryptoprocta*, while entitled to occupy a separate subfamily, Cryptoproctinae, "may be a very diverging root-form more or less allied to *Crossarchus* and *Herpestes*." Unquestionably there are a few characters, including those noted by Mivart (1882, pp. 193–195) and by Pocock (1916, p. 424), which would tend to support this view. In the details of the postcranial skeleton *Cryptoprocta* agrees with the Herpestinae in some characters, with the Viverrinae in others, notably the lesser reduction of the pollex and hallux, more retractile ungues, very long tail, long femur, slender lumbar parapophyses. In the lack of perineal glands and in the presence of circumanal glands opening into a prominent anal pouch, it agrees with the Herpestinae. Hence *Cryptoprocta* may be an offshoot from near the base of the viverrid stem, related on the one hand to the herpestines and on the other to the viverrines, hyaenas and cats. This leads again to the question whether *Cryptoprocta* should be classified with the Viverridae or the Hyaenidae or the Felidae, or perhaps set off in a family by itself.

The cats (Felidae) are indeed in many respects advanced beyond *Cryptoprocta* further than it is advanced beyond the most primitive known viverroids. The typical viverrines are slinking, long-bodied, very long-tailed robbers of small prey. The cats, on the other hand, while equally dependent on lurking and creeping forward toward their prey, are capable of making very sudden rushes and have shorter necks and tails and longer legs. The hands and feet, carrying to perfection the retractility of the claws that was introduced by the Viverridae, are more specialized in the loss of the hallux and generally in the greater compression of the digits. Several lines in the cat family have attained gigantic size (sabre-tooths, lions, tigers, and, to a lesser extent, others), together with further specialization of the dentition, chiefly by enlargement of the upper canines and varying emphasis of the carnassials. With the foregoing there has been a shortening of the muzzle, extreme widening across the zygomatic arches, reduction in the cheek teeth to $P\frac{2.3.4}{2.3.4}\ M\frac{1}{1}$, and other obvious but minor morphologic changes. The cats have no perineal scent glands and only small anal glands.

If we compare the entire skeletons of the modern *Genetta*, *Cryptoprocta*, *Felis*, there are so many peculiarities in common which distin-

guish them from those of other Carnivora that any suggestion of convergence seems like a mere *petitio principii*. In a general way the skeleton of *Cryptoprocta* is intermediate between those of *Genetta* and *Felis*, although much nearer the former in primitive characters. Thus, for example, according to Flower's table of the number of vertebrae ("Osteology," p. 86), *Cryptoprocta* has a very long proximally massive tail with 29 caudal vertebrae, as in *Genetta tigrina*, whereas in the Felidae the highest number listed (*Felis tigris*) is 25. Again, the anticlinal vertebra is D 11 in our mounted skeleton of *Cryptoprocta*, as it is in *Genetta* and *Civetictis*, whereas in the domestic cat it is D 10. In the cheetah, however, it is D 11. The cervical vertebrae of *Cryptoprocta* are relatively long, approaching those of the viverrids, while those of the typical cats are short. According to Pocock, the feet of *Cryptoprocta*, in regard to external characters, are "essentially Paradoxurine in type" . . . and they "differ, indeed, from the feet of the Felidae as profoundly as the feet of any Aeluroid differ therefrom." But a comparison of Pocock's figures of the plantar surfaces of the feet of *Cryptoprocta* with the feet of several preserved felids indicates that the differences while conspicuous are such as may distinguish a more primitive, semi-arboreal from a more specialized, semi-cursorial stage. Moreover, the skeletons of both manus and pes in *Cryptoprocta* are fundamentally viverrine in plan and differ from those of cats only in the lack of emphasis of certain features, such as the reduction of the pollex and absence of the hallux.

Broadly speaking, the dentition of *Cryptoprocta* has nearly reached the cat stage, while the auditory region of the skull remains on the viverrine plane.

These considerations, which could be supported by much detailed evidence, lead to the conclusion that *Cryptoprocta* is the little modified survivor of *Proailurus*, which was in turn at least a close relative of the direct ancestors of the cat tribe. *Cryptoprocta* therefore may be assumed to stand in a border zone between the Viverridae and the Felidae. But again, to which family shall it be assigned?

Here we are dealing not with a problem of phylogeny but with opposing practices of taxonomists. If we follow W. D. Matthew, who defended the so-called horizontal or group system of naming families, we shall place *Cryptoprocta* in the Viverridae and define the latter by the possession of certain relatively primitive characters common to *Cryptoprocta* and other viverroids. If, on the other hand, we follow the principle applied by Milne-Edwards, Filhol, and Trouessart (cited by Pocock, 1916, pp. 424, 425), we should refer *Cryptoprocta* to a primi-

tive subfamily of the Felidae. This procedure has been condemned by Pocock (*op. cit.*, p. 425) because it implies "an extension of the definition of the Felidae," with the result that the definition "ceases to be of scientific value." A third way is to follow the practice of certain modern mammalogists and make *Cryptoprocta* the type of a distinct family defined by its own peculiarities. This practice is generally approved and fashionable because it is supposed to be objective and without taint of hypothesis and speculation. However, the net result of this method is that in systematic mammalogy we are in danger of learning "more and more about less and less" until the classification of mammals has often degenerated into an increasingly large number of small families in

TABLE 3

Cryptoprocta-Felis

Skull Proportions	Index	Assumed Primitive *Genetta*	Range			Remarks
			Cryptoprocta	*Felis leo*	*Felis domesticus*	
Total skull length....		82	102	258	81	*Cryptoprocta* intermediate *F. dom.* reduced
Bizygomatic width..	I	50	61.8	82.6	80.2	*Cryptoprocta* intermediate
Max. cran. width ...	VI	54.7	65.6	81.4	77.4	*Cryptoprocta* intermediate
Forehead width.....	VIII	41.4	55	71.9	73.2	*Cryptoprocta* intermediate
Interlacr. width.....	II	61.5	67.5	67.4	79.3	*Cryptoprocta* relatively primitive *F. dom.* wide
Palatal width.......	X	73	79.6	111.0	127.3	*Cryptoprocta* relatively primitive
Lower dental arch width...........	XIV	54.1	63.8	85.7	100	*Cryptoprocta* relatively primitive
Bicondylar width.... Olfactory length.....	III	54.7	67.2	84.3	52.8	*Cryptoprocta* relatively primitive
Orbital length......	IV	34	36.1	45.8	50.9	*Cryptoprocta* very primitive *F. dom.* large
Basicranial length...	XI	47.2	42.6	48.3	43.4	Narrow range
Upper carnassial width......... ..		70.5	52.1	52.5	53.9	
Lower carnassial length....... ...		68.4	81.0	94.5	97.3	

which taxonomic grades of equal rank conceal widely different stages of phylogenetic relationship. In this instance it is perhaps more convenient to treat *Cryptoprocta* as a distinct section of the Felidae, intermediate between the Viverrinae and the Felidae and with possible relationships also to the Hyaenidae and Herpestidae.

Skull and dental measurements indicate that *Cryptoprocta* makes an excellent structural link between primitive Viverrinae and the Felidae.

The Ancestry of the Palm Civets and Their Allies

In his classic work on the Aeluroidea Mivart divided his subfamily Viverrinae into three sections:

A. Viverra, Viverricula, Fossa, Genetta, Prionodon, Poiana
B. Paradoxurus, Arctogale, Hemigalea, Arctictis, Nandinia
C. Cynogale

We have already considered the first section of Mivart's Viverrinae; the second section we regard as an essentially natural grouping despite the fact that it includes some extremely aberrant forms (*Arctictis, Nandinia*). The third section (*Cynogale*) is also plainly an offshoot from the first and not far from "*Hemigalea.*"

Mivart's classification was based not on any single characters but upon numerous characters, chiefly of the skull and dentition but by no means omitting coloration, the characters of the feet and the presence or absence of perineal and circumanal glands.

Pocock, on the other hand, using the presence or absence of the perineal scent glands as his leading criterion, splits up the Viverrinae into Viverrinae (*s.s.*) Fossinae, Prionodontinae, Paradoxurinae, Arctogalidinae, Hemigalinae, Cynogalinae and Nandiniidae. As a general result of our studies we conclude that Pocock's subfamilies consist of small natural groups; but that his system gives, as a whole, little or no hint of the broader relationships of these groups, and that the removal of *Nandinia* to form a family by itself is a retrogressive step which obscures the relatively close relationship of this form to the common stem of the hemigales and the palm civets.

Family Viverridae (*continued*)

Section *Paradoxurida*

Typically frugivorous offshoots of the Viverrida, Nandinia alone retaining the viverrine cheek teeth; p^4, m^1, *rarely enlarged, often much reduced in size, with blunt procyonine cusps;* m_1 *(except in Nandinia) with impaired shearing blades and expanded tuberculate talonid. Skull varying from*

carnivorous (Nandinia) to frugivorous (Arctictis) type. Feet secondarily plantigrade, finally becoming scansorial with wide plantar pads. Perineal glands typically well developed and surrounded by a muscular pouch, the perineal glandular area variable in position with reference to the vulva, which lies either in the middle (Paradoxurus) or in front (Arctictis) or nearly on the posterior wall (Arctogalidia) of the glandular area or pouch.

Nandiniinae

The "African palm civet" (*Nandinia*) presents a puzzling assemblage of features which long prevented us from perceiving what we now believe to be its true relationships with the stem of the Paradoxurinae. With regard to its perineal scent glands, Pocock (1915, pp. 409–411) has shown that *Nandinia* and *Arctictis* [of the Paradoxurinae] stand at the two extremes of specialization. For in the former the gland and its encircling pouch lie in front of the prepuce, while in the latter they lie behind it. Nevertheless, Pocock finds that " . . . the gland in the Paradoxurinae varies greatly in position in the genera discussed; but it is interesting to note the gradation in the situation it assumes, especially in the female." In short, the female scent gland of *Nandinia* approaches that of the female *Arctogalidia* of the subfamily Paradoxurinae in so far as the gland lies in front of the vulva. For this and other reasons Pocock in 1915 referred *Nandinia* to the Paradoxurinae.

In the same paper (pp. 390–393) he showed that except in minor details, the fore and hind feet of *Nandinia* closely resemble those of *Arctogalidia*, which was then regarded as a member of the Paradoxurinae. The same is true of the muzzle (p. 396) and the ear (p. 400). At that time (1915) Pocock did not hesitate to refer *Nandinia* to the Paradoxurinae, but in his Encyclopaedia Britannica article published in 1929 we find him isolating *Nandinia* as the type and sole occupant of a distinct family, Nandiniidae, characterized especially by certain structural features of the bulla and paroccipital region which will be discussed below.

One of the most conspicuous differences between *Nandinia* and the Paradoxurinae lies in the dentition. In *Nandinia* the carnassials (p^4 and m_1) although unusually small are of nearly pure "trenchant" type and directly derivable from those of the upper Eocene Viverrinae, whereas in the Paradoxurinae these teeth have acquired bluntly conical cusps of the frugivorous, procyonid type. Our experience with many similar cases in different groups of Carnivora suggests that in these features *Nandinia* is far more primitive and less specialized than any of the Paradoxurinae and that it belongs to a branch that antedates the acquirement of such frugivorous specializations in the latter subfamily.

The region of the middle ear and paroccipital in *Nandinia* is so strikingly different from that of most other modern Viverridae that Pocock (1929, p. 898) selects these as the leading characters in defining

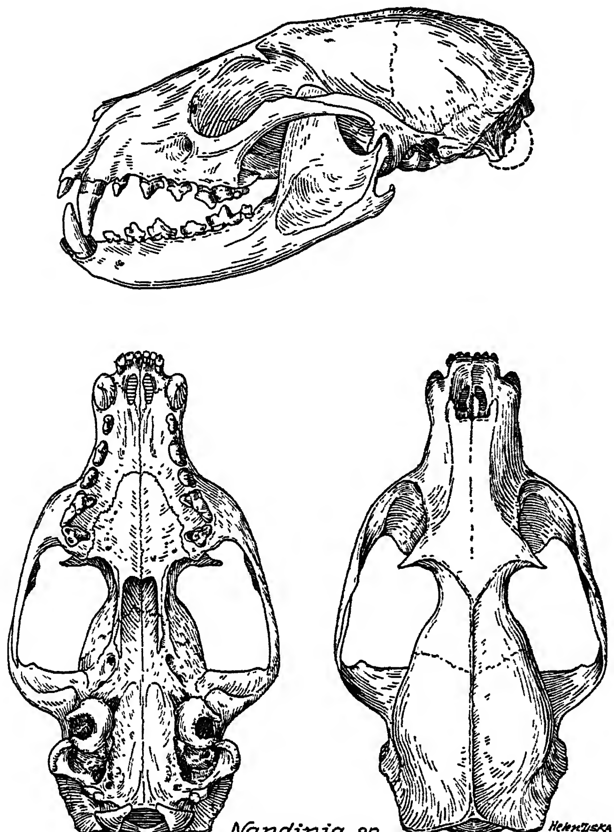


FIG. 14. *Nandinia*, three views of skull. × ⅔.

the "family" Nandiniidae. Nevertheless, further analysis and comparison convince us that *Nandinia's* claims to relationship with the Paradoxurinae, which Pocock formerly admitted on the basis of the characters of the scent glands and feet, can be even more securely based upon those of its auditory and basicranial region. In the dried skull of *Nandinia* the entotympanic portion of the bulla is normally absent and, as Teilhard de Chardin has noted, this condition is apparently found also in the fossil skull of the upper Eocene *Palaeoprionodon Lamandini*. But in other upper Eocene carnivores (*e.g., Cynodon typicus, Plesictis robustus* (de Chardin, *op. cit.*, pp. 52, 58)) the entire bulla (in these cases possibly including only the ectotympanic) is well ossified and from the

inflation of the paroccipital region in *Palaeoprionodon Lamandini* it seems evident that in this genus the entotympanic bulla was also inflated. Possibly its bony wall was too fragile to be preserved in the fossil or possibly a failure to ossify had already set in. In either case inspection of a number of dried *Nandinia* skulls, including a very immature one, suggests that the failure to ossify is not a primitive but a highly specialized feature peculiar to this genus alone among modern Viverridae.

The basis cranii of *Nandinia* is remarkably short and wide, quite different from the long, narrow, basioccipital-basisphenoid region of typical Viverrinae. But among all the viverroid Carnivora the only ones that approach or even exceed *Nandinia* in this region are the genera *Arctogalidia* and *Arctictis* of the subfamily Paradoxurinae.

According to Pocock (1929, p. 898), the family Nandiniidae "differs from the rest of the Aeluroidea," among other characters, "in the large size and shelf-like form of the mastoid portion of the skull and the backward direction of the paroccipital away from the bulla, two characters in which it resembles many of the typical Arctoidea." But here again Pocock seems to have overlooked the remarkably close approach of *Nandinia* to *Arctictis* of the Paradoxurinae, not only in these features but in the entire occiput and even in other parts of the skull. Assuredly, the resemblances mentioned above in the mastoid and paroccipital to the Arctoidea are worthless as indicative of phylogenetic relationship with the Arctoidea, for the reason that *Nandinia* as a whole is a true aeluroid with only remote relationship with the "typical Arctoidea," which are the dogs and bears.

In its external appearance, as figured by Lang (in Allen's report on the Carnivora collected by the American Museum Congo Expedition, Pls. XX, XXI), *Nandinia* is presumably fairly primitive, that is, not unlike such Viverrinae as *Poiana*, in retaining longitudinal rows of spots on the sides and remnants of rings on the tail, whereas some of the Paradoxurinae have acquired a more uniform coloration; others, however, retain traces of the longitudinal rows of stripes on the sides and more or less distinct dark rings on the tail.

Thus the logic of facts forces us to the conclusions (1) that *Nandinia* is an offshoot from near the base of the paradoxurine stem; (2) that it has retained rather primitive trenchant cheek teeth; (3) that these have, however, suffered from that genetic size-reduction factor which has gone to much greater lengths in certain of its allies among the Paradoxurinae; (4) that the very characters of the mastoid and auditory regions which were supposed to be evidences of exceptional antiquity

are rather new specializations which may be closely matched (except for the non-ossification of the entotympanic) among the most specialized of the Paradoxurinae; (5) that the removal of *Nandinia* to a distinct family, however much it may facilitate the construction of definitions based on arbitrarily chosen characters, yet has the unfortunate effect of obscuring the close relationship of the African *Nandinia* to its Oriental allies, the palm civets.

Arctogalidiinae

The "small-toothed palm civets" (*Arctogale, Arctogalidia*) of southeastern Asia and the Sondaic peninsula were classified by Mivart (1882, pp. 143, 163) with the true palm civets (*Paradoxurus*), the binturong (*Arctictis*), as well as with "*Hemigalea*" and *Nandinia*. Externally they much resemble *Paradoxurus* in general shape and in the length of the legs and likewise have very long tails. They often have three dark bands running along the back and in some species the short fur is dusky or brown on the back.

They have been separated by Pocock (1933, pp. 969, 977) as a distinct subfamily, Arctogalidiinae, chiefly on the basis of certain characters of the perineal scent glands, which may now be discussed. The first character of the Arctogalidiinae is the "absence of the perfume-gland in the male." Now this gland is present in males of the subfamilies Viverrinae, Paradoxurinae, Hemigalinae, Cynogalinae (Pocock, 1933, p. 977), to all of which, according to our own studies, the subfamily Arctogalidiinae is more or less closely related. Hence the absence of the gland in males of the Arctogalidiinae may possibly be due to its loss. The absence of the perineal scent gland, at least in males, is characteristic, according to Pocock's data (1929, p. 898) of various other genera and subfamilies (Galidictinae, Fossinae, Euplerinae, Prionodontinae, *Cryptoprocta*). Whether or not the absence of this gland in males is a specialized character in all these subfamilies, its absence in the male *Arctogalidia* no more separates that genus phylogenetically from *Paradoxurus* than it allies *Arctogalidia* with any of the other forms that likewise lack the gland. At any rate, these glands are but poorly developed in the male *Paradoxurus hermaphroditus* (Pocock, 1915, p. 405, Fig. 7C) in comparison with their rich development in the male *Paradoxurus larvatus*. A further reduction would result in "no glands," as in the male *Arctogalidia*.

As the scent gland is absent in the male *Arctogalidia*, it is perhaps not surprising that the space where it was, or where it failed to develop, should be covered with hair and not "bald" as it is in the male *Paradoxurus* (Pocock, 1915, p. 405). Moreover the forward displacement

of the prepuce far in advance of the scrotum in *Paradoxurus* (*Paguma*) *larvatus* and *Arctictis binturong* may have afforded plenty of space for the exceptional development of their glands. On the other hand, the enfeeblement of these glands in the particular *Paradoxurus hermaphroditus* figured by Pocock (p. 405) did not affect the position of the prepuce. So, we may well suppose, the complete loss of the glands in the male *Arctogalidia* left the prepuce far forward but gave the body hair an opportunity to invade the area between the prepuce and the scrotum.

Turning to the conditions of the perineal region in the female *Arctogalidia*, we note that Pocock (1933, p. 977) assigns a subfamily character to "the position of the gland in the female in front of the vulva and clitoris, where it is represented externally by two low ridges of naked skin capable of being folded over in front of the generative orifice and continuous behind on each side with the naked skin surrounding it." Inspection of Pocock's excellent figures of 1915 shows that in *Arctogalidia* the glandular area is chiefly in the middle of a depression surrounded by an upstanding flap of skin, whereas in *Paradoxurus* (*Paguma*) *larvatus* and *Arctictis binturong* the glands are paired and highly developed so that their labia are well filled with glands. But these differences seem to be essentially in degree rather than in kind.

As to the relative position of the gland or its pouch and the vulva, the latter lies near the center of the pouch in *Paradoxurus hermaphroditus* but in *Arctogalidia* the vulva is immediately in front of the rear wall of the pouch. This is one of the leading diagnostic characters assigned for the subfamily Arctogalidiinae. In this respect *Arctogalidia* may well be specialized in the opposite direction from *Arctictis*, in which the vulva lies wholly in front of the pouch. Possibly the primitive position of the pouch was essentially as it is in *Galidia* and *Galidictis* (Pocock, 1915, *Ann. Mag. Nat. Hist.*, XVI, Pl. XIV), namely, surrounding the vulva but with the latter near the front border of the pouch.

The feet of *Arctogalidia*, according to Pocock's figures (1915, *Proc. Zool. Soc. Lond.*, p. 391) are assuredly intermediate in general structure between the narrow digitigrade feet of the typical Viverrinae and the very wide pad-like scansorial feet of *Arctictis* (p. 390). In the hind foot the pads of the third and fourth digits are quite distinct, whereas in the Paradoxurinae these pads tend to fuse at the base, especially in *Arctictis*.

The dorsal view of the skull of *Arctogalidia* much resembles that of *Paradoxurus* except for minor differences noted by Pocock (1933, p. 977). In the region of the bulla Pocock notes that in *Arctogalidia* the tympanic (ectotympanic) early fuses with the bulla (entotympanic), whereas in the Paradoxurinae these remain either suturally distinct or separated

by a prominent groove. Moreover the entotympanic bulla lacks the prominent ventral ridge which is conspicuous in the Paradoxurinae, while the paroccipital process is not prolonged downward so acutely and

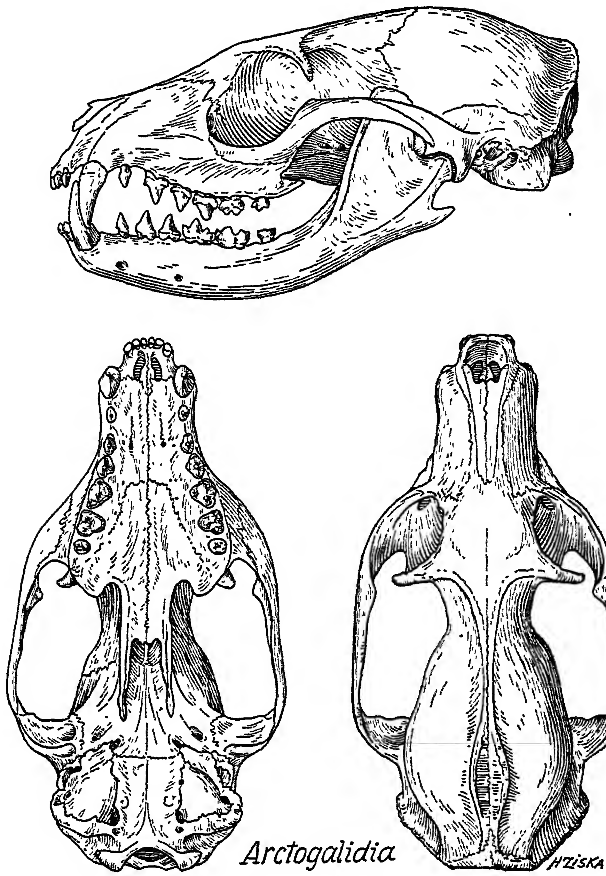


FIG. 15. *Arctogalidia*, three views of skull. × ⅔.
P^4 has a small but distinct metastyle not well figured in the drawing.

is wider than in the Paradoxurinae. The latter character may possibly be associated with the less vigorous development of the nuchal, temporal and digastric muscles and with the more slender jaw. The tubular restriction and posterior prolongation of the palate may be conditioned by some peculiar muscular development of the internal pterygoid muscles; this is apparently also indicated by a heavy and sharp horizontal torus on the inner side of the middle third of the mandible.

By far the most conspicuous feature of *Arctogalidia* is the reduction in size of its cheek teeth and the tendency of p^4, m^1 and the functionally

associated lower teeth to lose the carnassial character of primitive viverrids. In the light of many studies on the evolution of the mammalian dentition we can see in these simple cheek teeth of *Arctogalidia* but little of the "primitive trituberculy" and many of the stigmata of degenerative specialization, which have produced a close convergence to the conditions in *Arctictis*.

TABLE 4

Skull Proportions	Index	*Genetta*	*Fossa*	*Nandinia* (4)	*Arctogalidia*	*Arctictis*
Total skull length		82.0	93	87.5–95.2	99.0	120
Bizygomatic width	I	50.0	53.4	57.8–69.3	54.0	62.4
Max. cran. width	VI	54.7	60.7	62.7–67.4	62.5	75.0
Forehead width	VIII	41.4	38.1	38.8–45.7	52.0	66.7
Interlacr. width	II	61.5	58.1	52.1–70.1	55.6	80.0
Palatal width	X	73.0	60.2	68.0–81.9	61.4	82.0
Lower dent. arch width	XIV	54.1	48.0	55.3–67.2	55.6	65.9
Lower jaw spread	XII	49.1	47.0	44.6–53.5	45.7	47.3
Bicondylar width				20.3–22.0		
Olfactory length	III	54.7	76.4	65.8–74.7	73.7	81.8
Orbital length	IV	34.0	35.6	34.1–37.8	35.1	30.3
Basicranial length	XI	47.2	42.1	36.6–45.9	37.9	42.4
Upper carnassial width		70.5	78.8	64.8–71.4	93.1–115.7	109.1
Lower carnassial length		68.4	67.1	76.1–83.6	61.8– 60.3	61.1

TABLE 5

Skull Proportions	Index	*Genetta*	*Arctogalidia*	*Arctictis*	*Paradoxurus*
Total skull length		82.0	99.0	120.0	107.2
Bizygomatic width	I	50.0	54.0	62.4	59.7
Max. cran. width	VI	54.7	62.5	75.0	61.9
Forehead width	VIII	41.4	52.0	66.7	—
Interlacr. width	II	61.5	55.6	80.0	55.8
Palatal width	X	73.0	61.4	82.0	73.5
Lower dental arch width	XIV	54.1	55.6	65.9	50.2
Lower jaw spread	XII	49.1	45.7	47.3	45.7
Bicondylar width					
Olfactory length	III	54.7	73.7	81.8	72.0
Orbital length	IV	34.0	35.1	30.3	31.9
Basicranial length	XI	47.2	37.9	42.4	42.5
Upper carnassial width		70.5	93.1–115.7	109.1	90.0
Lower carnassial length		68.4	61.8– 60.3	61.1	68.8

The skulls of this subfamily are of intermediate size and more or less "mesocephalic." They show a wide range of variability in the transverse measurements of the forehead, interlacrymal diameter, palate and lower dental arch. All the measurements appear to be consistent with the derivation of the subfamily from the primitive viverrine stock. The cheek teeth have reduced the carnassial character and become blunt cusped; in some genera they are also very small.

Paradoxurinae

Within the subfamily Paradoxurinae (*s.s.*) Mivart (1882, pp. 160–165) notes a wide range of variation, especially with regard to coloration and to the size and degree of bluntness of the cheek teeth. The second upper and lower molars are entirely wanting in two skulls recorded by Mivart. This tendency toward the elimination of m^2 occurs elsewhere in the family, especially among the Viverrinae, Priodontinae and Cryptoproctinae. In all the genera there is a more or less pronounced tendency toward the reduction of the metastyle blade of p^4 and the consequent transformation of a trenchant into a bicuspid or tricuspid premolariform crown. The fourth upper premolar and the first upper molar tend to

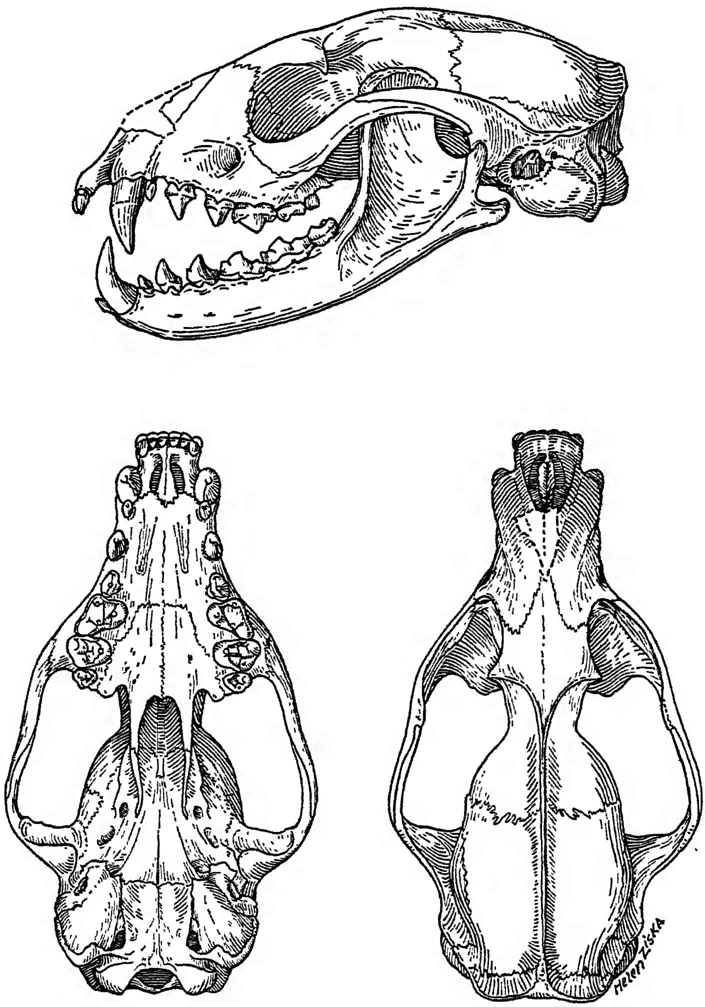


FIG. 16. *Paradoxurus*, three views of skull. × ⅔.

become small in *Arctictis*, but they increase in size in *Paradoxurus* and *Paguma*. The second upper molar crown is reduced to a minute rounded oval in *Arctictis*. These retrogressive changes in the dentition have conditioned, or have been conditioned by, a change from a prevailingly carnivorous to a more or less frugivorous regimen.

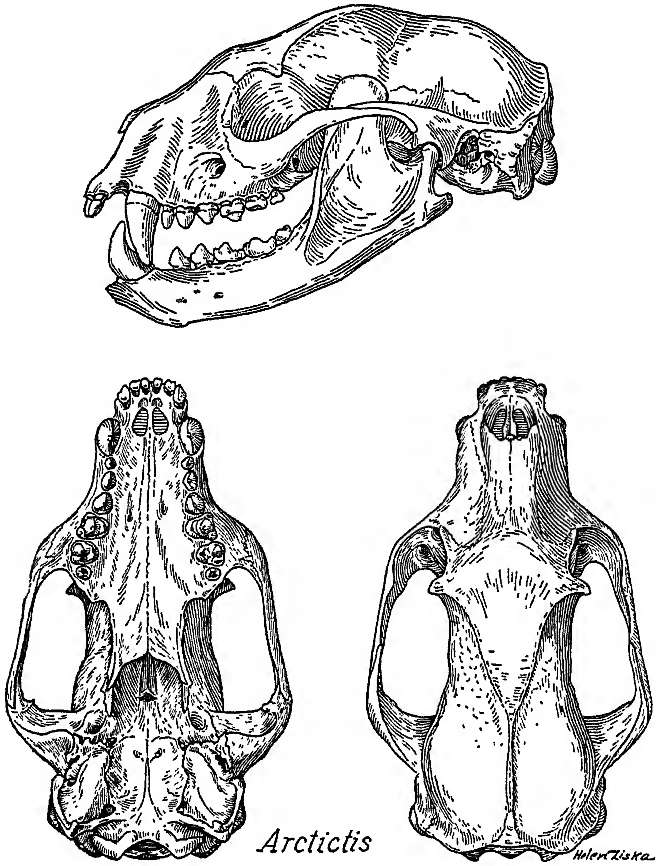


FIG. 17. *Arctictis*, three views of skull. × ½.

The auditory bulla of *Paradoxurus* is described by Flower (quoted by Mivart, p. 161) as "conical, broad, and truncated behind, pointed in front and rather compressed on the sides, which meet in a ridge." Closely related forms of bullae are seen in the other genera of the subfamily. The basi-occipital region becomes remarkably short and wide in *Arctictis* and it has been shown above that the backwardly inclined paroccipital process of the same genus indicates how the conditions of this region in *Nandinia* have probably been derived from those seen in the Paradoxurinae.

The hands and feet of the Paradoxurinae (as described by Pocock, 1915, pp. 387–391) are adapted to arboreal habits. They seem to be readily derivable from those of typical Viverrinae (*cf.* Allen, 1924, p. 123) by transverse spreading, by the probably secondary enlargement of the hallux, and by the marked expansion of the interdigital, metacarpal and metatarsal pads, together with the loss or reduction of hairy regions, which in the more primitive *Genetta* restrict these pads transversely.

In short, the Paradoxurinae as a whole are distinguished by the high specialization and variability of their perineal glands and by their tendency to acquire more or less frugivorous modifications of the cheek teeth, often but not always associated with reduction in size. The entotympanic bullae are conical, pointed in front, broad behind, often ridged. The basi-occipital region tends to be short and wide and the palate much produced backward beneath the narrow pterygoid flanges of the alisphenoid. The pentadactyl hands and feet are secondarily widened, with spreading digits and large areas for contact with the limbs and branches of trees among which they live. This and the occasional tendency for the tail to be coiled without being actually prehensile may be regarded as secondarily arboreal adaptations. The derivation of the subfamily seems to be from the primitive viverrine stem but they have become specialized in connection with arboreal life. In this connection it is hardly necessary to assume with Matthew (1915) that arboreal characters in modern Carnivora (as among the Procyonidae) are due to direct inheritance from hypothetical arboreal pro-placentals or even from arboreal Miacidae. The comparative morphological evidence, as we read it, would suggest that the plantigrade, frugivorous Paradoxurinae represent a secondarily arboreal specialization from a digitigrade, terrestrial, predaceous stem of the viverrine stock.

The Ancestry of the Fanaloka (*Fossa*), the Hemigalines and Related Forms

Family Viverridae (*continued*)

Section *Hemigalida*

Carnivorous (Fossa), omnivorous (Hemigalus), piscivorous (Cynogale), or myrmecophilic (Chrotogale, Eupleres) derivatives of the Viverrida; cheek teeth less carnassial than in stem Viverrinae, showing progressive stages either toward the procyonine, blunt-cusped type (Fossa, Hemigalus, Cynogale), or, by decrease in transverse width and increase in length, toward a small-toothed, secondarily sharp-cusped myrmecophilic stage (Chrotogale, Eupleres). Bony muzzle long and low, forehead narrow, postorbital con-

striction moderate to extreme (Cynogale) or secondarily widened (Eupleres). Ectotympanic but little or not at all inflated, enclosing the auditory meatus in a prominent oval, directed obliquely upward and backward; elongation of entotympanic portion of bulla variable (slight in Fossa, extreme in Cynogale). Feet of viverrine derivation, primarily subcursorial; pes with moderate (Hemigalus) to very long metatarsus (Fossa); manus somewhat widened, becoming partly fossorial (Eupleres). Perineal scent glands, if present, not large, tending to reduction, especially in males (Hemigalus), or absent (Fossa, Eupleres).

Fossinae

One of the clues to the ancestry of the Hemigalinae is to be found, we conclude, in a study of the Madagascan genus *Fossa*. This animal has often been confused with the true fossa (*Cryptoprocta*), which has a much more cat-like skull and dentition. It is called the "Fossane" (*Fossa Daubentoni*) by Beddard but A. L. Rand (1935, p. 93) states that this "striped civet of the humid forest" is known as "fanaloka" to the natives. In three stomachs of this species he found insect matter, in another a lizard 100 mm. long. Hence the animal is insectivorous-carnivorous.

Its skull and dentition are in many respects intermediate between the Viverrinae (*s.s.*) and the *Hemigalus* group. The skull as viewed from above is similar to that of *Genetta* but the muzzle is considerably longer, the nasals long and not tapering posteriorly, the median groove between opposite nasals and frontals much more pronounced. In this top view of the skull, save that the median dorsal groove extends back between the frontals, *Fossa* approaches *Hemigalus*.

The posterior chamber of the bulla of *Fossa* is relatively short (anteroposteriorly), wide and sub-globose posteriorly, whereas in *Hemigalus* (and still more in the Viverrinae) it is long and narrow. The globose bulla of *Fossa* is enclosed posteriorly by the thin concave shell formed by the paroccipital region much as in *Hemigalus*, except that it lacks the downward extension and transverse ridge of the latter genus.

The upper premolars of *Fossa* are, on the whole, not dissimilar from those of *Hemigalus*, save that p^4 is larger and has a better developed metastyle shear, while m^1, m^2 are more primitive, that is, with sharper cutting blades and wider transverse diameters. M^1 has a secondary metastyle. In the lower cheek teeth, m_1 in *Fossa* has a large sharp, two-bladed trigonid and a low, relatively small talonid, whereas in *Hemigalus* the blades of the trigonid are small and closely appressed to each other while the talonid is expanded and well cuspidate.

The cheek teeth of *Fossa* are of primitive generalized carnivorous type, much like those of Eocene Miacidae, and far more primitive than those of the Paradoxurinae and Arctogalidiinae.

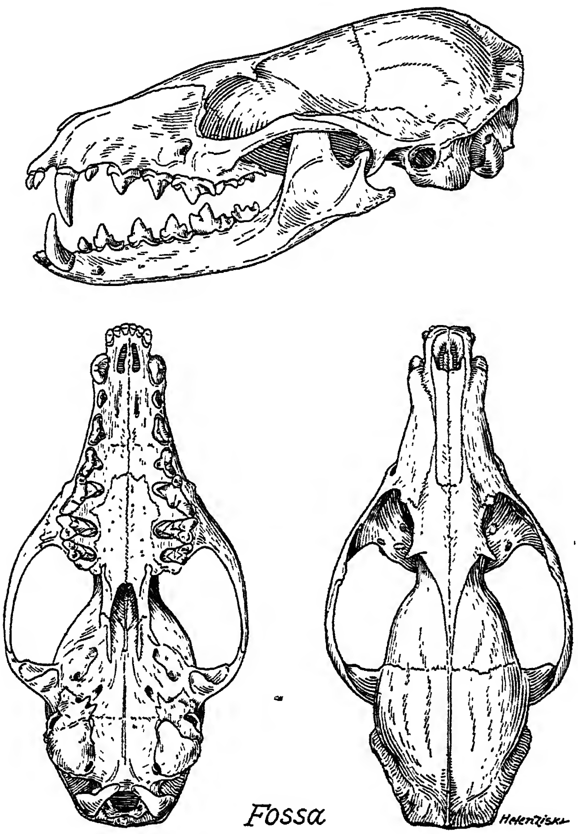


FIG. 18. *Fossa*, three views of skull. × ⅔.

The skull of *Fossa* is far less specialized than those of the Paradoxurinae: its muzzle less thickened, its palate less spreading and less prolonged posteriorly, its basis cranii less widened, its entotympanic bullae not laterally compressed or ventrally keeled, its paroccipital processes not at all posteriorly ridged, its occiput less widened, etc. The skull is remarkably like that of *Hemigalus*, except in detail (more slender muzzle, more globular entotympanic bullae). The dentition is also more primitive.

The hind feet of *Fossa*, according to Mivart (1882, pp. 150, 151; see also Pocock, 1915, Pl. XIII, Fig. 6), have a greatly prolonged metatarsal segment, with a much reduced hallux placed high above the other digits;

the claws are slightly, if at all, retractile and the general effect, as Pocock notes, is "canine" rather than "feline." Nevertheless this specialized foot is apparently derivable from the type exemplified in *Genetta*, especially *Genetta servalina* (Allen, 1924, p. 129, Fig. 19), and the same is even more true of the manus (*cf.* Pocock, *op. cit.*, Pl. XIII, Fig. 5), which is much shorter than the pes and not widely different except in details from the manus of *Genetta servalina* (Allen, 1924, p. 129, Fig. 19).

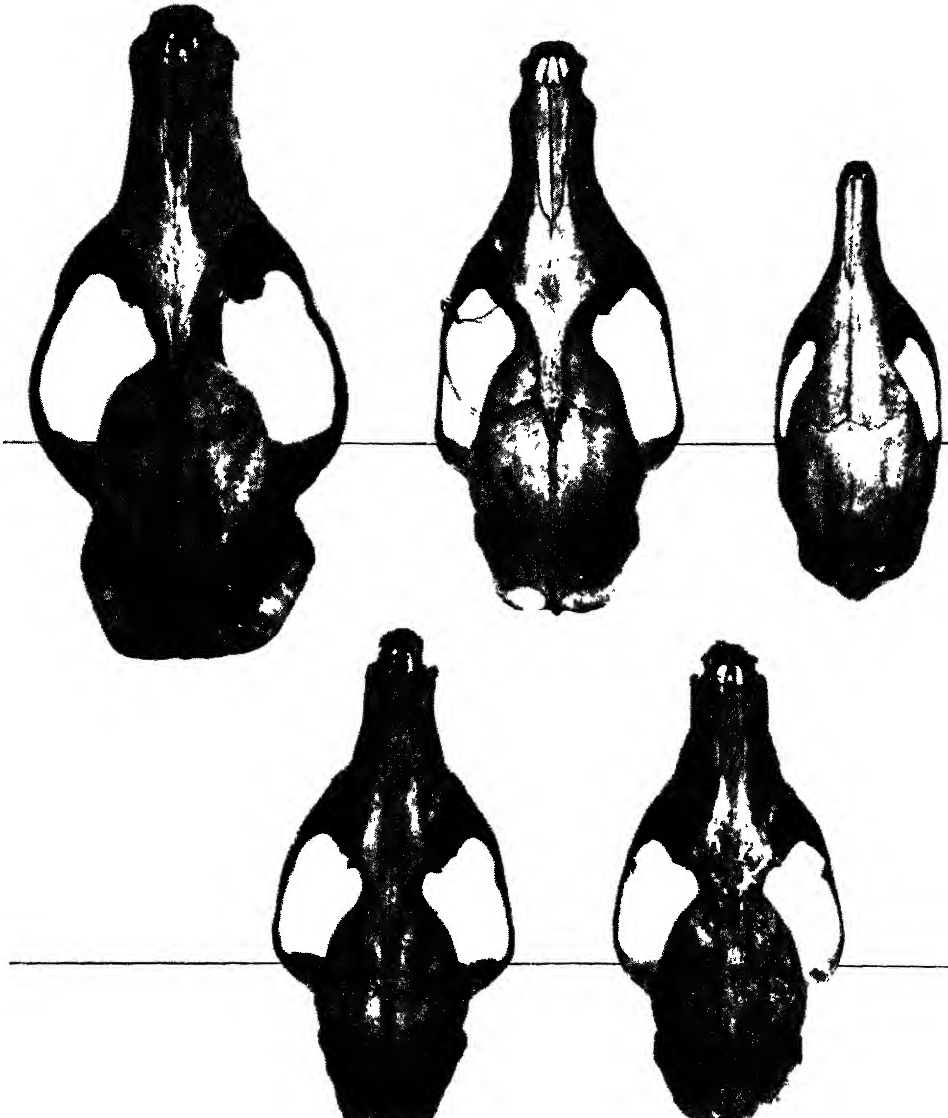
The manus of *Fossa* (Pocock, *op. cit.*, Pl. XIII, Fig. 5) is much shorter and apparently less specialized than the pes. It differs from that of *Genetta servalina* chiefly in the reduction of the pollex and in the invasion of the hair between the interdigital and hypothenar pads.

Thus with regard to its pes *Fossa* seems to represent a somewhat specialized cursorial derivative of the primitive viverrine pes, while in its manus it is less specialized away from the primitive viverrine type.

Fossa, according to Pocock (1915, p. 346), has no perineal scent glands (1915, p. 344; 1929, p. 898) and he therefore erects for it a separate subfamily, Fossinae. But in *Hemigalus*, which, as will presently be shown, seems also to be related to the stem of the Paradoxurinae, the perineal scent gland, according to Pocock (1915, pp. 155, 156; 1929, p. 898), is reduced in both sexes, and in *Arctogalidia* of the Arctogalidiinae, according to the same authority, this gland is absent in the male. The complete absence in *Fossa* may be due either to the loss of this organ or to its possible absence in the common ancestors of the Viverrinae, Paradoxurinae, and Herpestinae. In either case its absence is not valid evidence against the inferences here drawn, which are: (1) that *Fossa* is a side branch from the common stem leading to the Paradoxurinae and the Hemigalinae; (2) that it is much more primitive than either of the latter in the chief characters of its dentition and skull; (3) that it is somewhat aberrantly specialized in the elongation of the tarsus and reduction of the hallux.

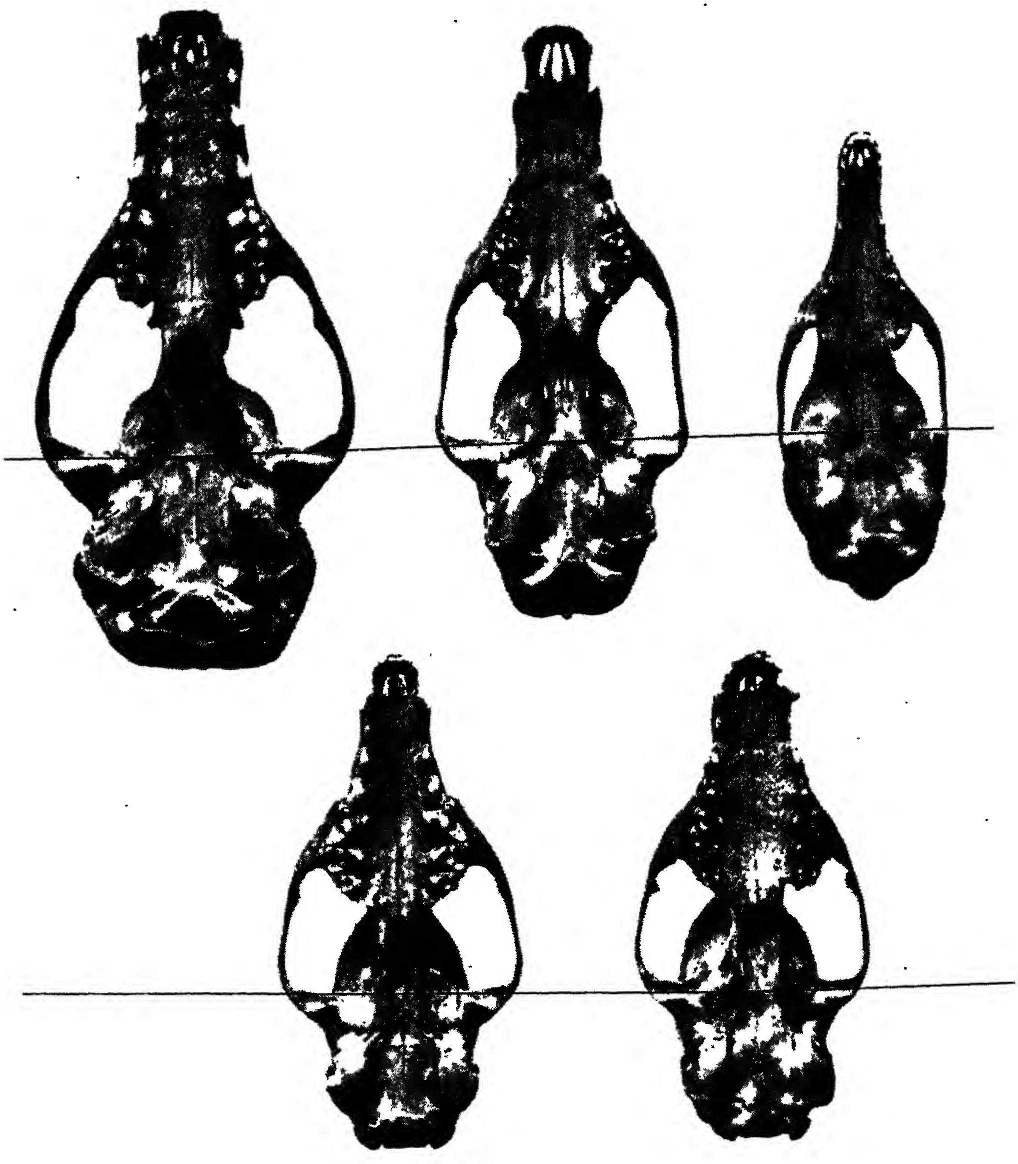
To sum up, in view of the foregoing considerations it seems convenient to treat *Fossa* as the sole representative of the subfamily Fossinae of Pocock. Its lack of perineal glands distinguishes it from the Viverrinae, Paradoxurinae and Hemigalinae; its tarsus is much more elongate than that of any other viverrid; its sub-globose bulla is more primitive than the anteroposteriorly elongate bulla of the Viverrinae; its sharp-cusped cheek teeth are only less carnassial in form than those of the most primitive lower Oligocene viverrids, and its skull as a whole is close to the latter.

PLATE III



SKULLS OF THE HEMIGALID SERIES. Top view. × ⅔. Upper row, left to right: *Cynogale, Chrotogale, Eupleres.* Lower row, left to right: *Fossa, Hemigalus.*

PLATE IV



SKULLS OF THE HEMIGALID SERIES. Underside view. × ⅔. Upper row, left to right: *Cynogale, Chrotogale, Eupleres.* Lower row, left to right: *Fossa, Hemigalus.*

Hemigalinae

The subfamily Hemigalinae (in the restricted sense) includes several genera of Asiatic Viverridae in which the presumably primitive, essentially viverrine fur pattern of longitudinally arranged blotches has been changed into more or less well defined sharp transverse stripes, especially on the shoulders and back. The skull of *Hemigalus* on the whole differs but little from the primitive viverrine type, the most outstanding difference being in the much less carnassial character of the cheek teeth. For in this subfamily there is a pronounced tendency for p^4 and m^1 to become alike in supporting conical pointed tubercles with reduced shearing blades. Correlated with this, the outer carnassial angle has been greatly opened out, while the inner carnassial angle is markedly reduced. There has been also a decided tendency to develop a prominent fossa in the middle of the crown of both p^4 and m^1; in m^1 there is a marked upgrowth of the protocone-protoconule ridge in front and of the protocone-metaconule ridge behind. The sharp crests of these ridges bear several small cuspules. M^2 has increased in size, added a hypocone and become rounded quadrangular. The lower carnassial has lost its blade-like character, the paraconid and metaconid being twinned, the hypoconid and basin of the talonid are enlarged, m_2 is secondarily increased in size.

The deciduous upper cheek teeth (pd^3, pd^4) of *Hemigalus*, as described by Leche (Zool. Jahrb., Bd. XXXVIII, pp. 296–299), are delicate, compressed, elongate, with reduced inner cusps and accessory cuspules on the lingual edges; dp^3 is much less carnassial and dp^4 less molariform than is the case in the Viverrinae.

Likewise in the lower permanent teeth m_1 has lost most of its primitive sectorial character, the paraconid and metaconid being squeezed nearly together, the talonid much enlarged, lowered nearly to the level of the hypoconid, and cuspidate; m_2 is secondarily enlarged, with basin-like cuspidate crown. As to the deciduous cheek teeth (dp_3, dp_4), as described by Leche (*ibid.*, p. 297), the former is much compressed and serrately cuspidate, the latter bears a pronounced carnassial trigonid and a large cuspidate talonid.

To Oldfield Thomas (1927, p. 48) and Mr. Hinton, as well as to Forsyth Major, the cuspidate character of the permanent molars of the Hemigalinae was a sign of primitiveness, chiefly because the molars of *Hemigalus* and its allies are more "multituberculate," using this word merely in its etymological sense, and because the multituberculates were possibly the oldest known main branch of the mammals. To us, on the other hand, such a dentition as that of *Hemigalus* illustrates in the cheek

teeth a fairly advanced stage of the process that one of us (W. K. G.) has named "secondary polyisomerism," wherein there is a tendency toward the loss or softening down of the originally sharp contrasts between adjacent members of the cheek teeth series, especially between p^4 and m^1.

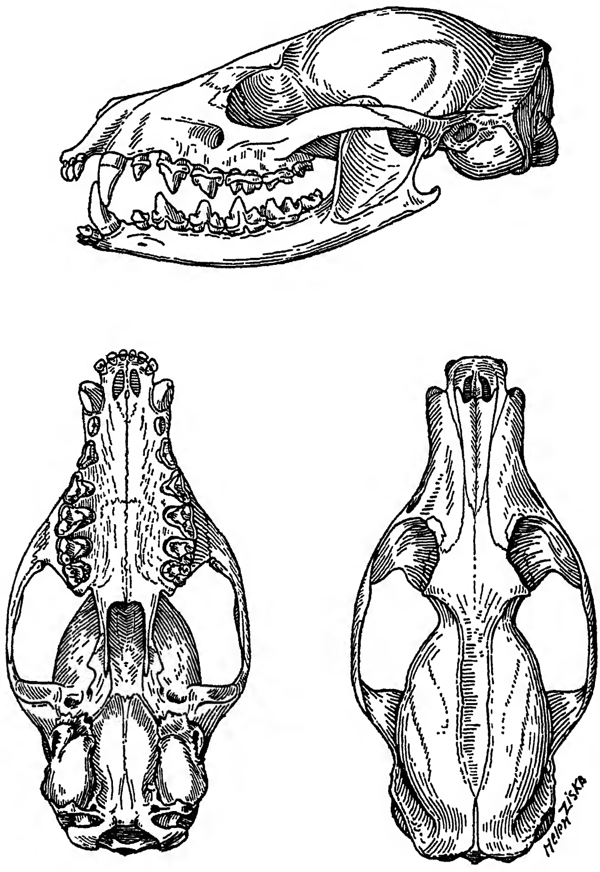


FIG. 19. *Hemigalus*, three views of skull. × ⅔.

The skull of *Hemigalus* retains much of the primitive viverrine stamp illustrated in the recent *Fossa* and in the Upper Eocene *Palaeoprionodon*. As compared with the typical Viverrinae, *Hemigalus* has the bulla smaller, with the entotympanic less markedly inflated and less antero-posteriorly elongated. The perineal scent glands of *Hemigalus*, according to Pocock, are somewhat reduced; perhaps they have never been developed to the same extent as in the typical Viverrinae.

The pes is shorter than in *Genetta*, the metatarsal pads having disappeared; the manus, however, is nearer to that of *Genetta*, and on the

whole the feet of *Hemigalus* are intermediate in stage between those of the semidigitigrade Viverrinae and those of the secondarily plantigrade Paradoxurinae.

The skull of the closely allied *Diplogale* (Oldfield Thomas, 1892, p. 222) of Borneo is somewhat more slender than that of *Hemigalus*, with more delicate teeth.

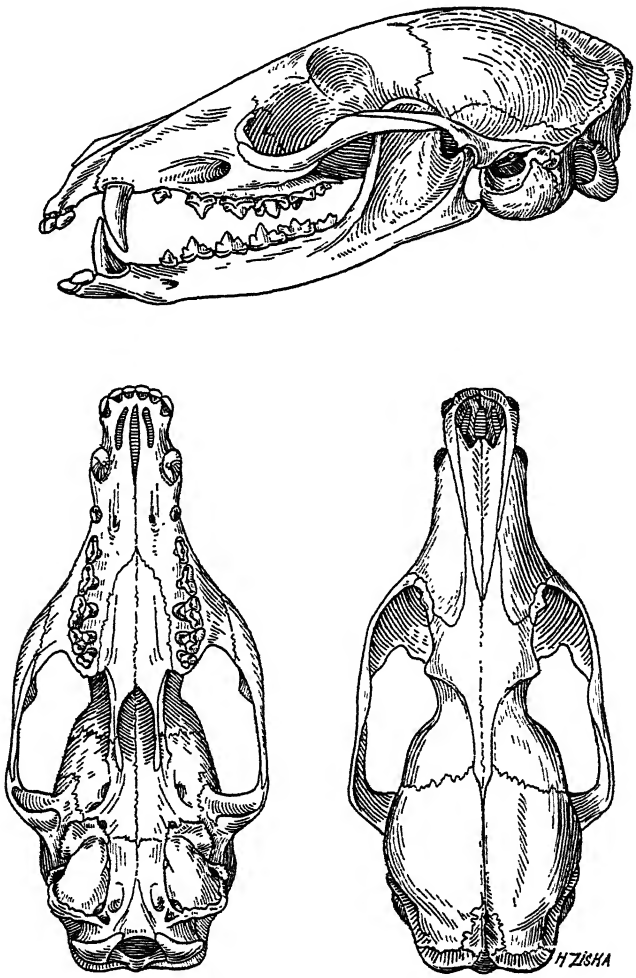


FIG. 20. *Chrotogale*, three views of skull. × 2/3.

The strange *Chrotogale owstoni* of Southeastern Asia, originally described by Oldfield Thomas, is a bizarre and specialized derivative of *Hemigalus*, to which it is closely allied by its color pattern (Pocock, 1933, p. 1011 and Pl. II). Through the kindness of Dr. W. H. Osgood of the

Field Museum, Chicago, a skull of an adult male collected by the Kelley-Roosevelts Expedition in Tonkin has been loaned to us for study. The skull as a whole is about 4½ inches long, distinctly larger than that of *Hemigalus hardwickii*. It is relatively very long, superficially suggesting the skull of a borzoi hound, with its long pre-orbital face. Owing to the lengthening of the entire skull, especially distally, the height of the mandibular condyle above the level of the cheek teeth is markedly reduced as compared with that of *Hemigalus*. The distal end of the muzzle has grown downward and the lower incisors have become sharply procumbent. The upper and lower canines are exceptionally long, gently recurved, with delicate, very sharp points. The premaxillae are produced forward, leaving a relatively long diastema behind the lateral incisor for the accommodation of the vertical lower canines. The upper incisors form a close-set cutting edge not unlike the incisive arch of a horse or that of a kangaroo; the incisors increase rapidly along their "mesiodistal" edges from i^1 to i^3, the latter being exceptionally long. The lower incisors are procumbent in a spatulate arch, the lateral incisors (i_3) being anteroposteriorly elongate; their crowns are convex on the labial side and short vertically.

The vertical interlocking of the upper and lower canines permits only a vertical and slightly posterior movement of the sharp edges of the lower incisors, which glide along the palatal surfaces of the upper incisors. Such an arrangement would hardly appear suitable for the shearing of blades of grass and the short crowns would not stand up against the abrasive reaction of silicious stems. Willoughby Lowe of the Kelley-Roosevelts Expedition noted that in the two specimens secured by him the stomachs contained earthworms (Osgood, 1932, p. 257), so that the peculiar form and arrangement of the incisors may be of value in cutting earthworms into convenient lengths.

Behind the upper incisor arch the anterior palatal foramina form a pair of prominent longitudinal slits 8.4 mm. long and about 1 mm. wide; between them is a conspicuous median slit or septal foramen 14 mm. long and 2 mm. wide. This system of slits could readily be derived from the conditions in *Hemigalus* by the elongation of the premaxillae and the great emphasis of the median or septal foramina.

Behind the canines the entire dental and mandibular arches show an early stage in that enfeeblement of the dentition which has been carried further in such myrmecophilic forms as the marsupial *Myrmecobius* and the insectivore *Hemicentetes*. As in those genera, the palatal and dental arches of *Chrotogale* are long and narrow, the individual cheek teeth laterally compressed, enfeebled and elongate, especially p^2, p^3

and $p_2 - m_1$. In these and other features *Chrotogale* is nearer to *Diplogale* than to *Hemigalus*. All the cusps are more slender and delicate than in *Hemigalus* but the difference from *Diplogale* is the lesser of the two. The cheek teeth lack the cuspidate crests, conical protocones and internal cingula of *Hemigalus* and p^4 looks a little more like a normal carnassial. Those who accept certain current applications of Dollo's "law of the irreversibility of evolution" would doubtless be unwilling to admit that a more simple, less complicated tooth form may have evolved out of a more complex earlier pattern. Nevertheless we suspect that the simpler pattern of the cheek teeth of *Chrotogale* as compared with those of *Diplogale* is the result not so much of divergent evolution from a remote common ancestor as of the secondary reduction in *Chrotogale* of the transverse diameter and increase of anteroposterior measurements of $p^2 - m^1$.

The lower cheek teeth of *Chrotogale* afford strong additional evidence that this genus is far more nearly allied to *Diplogale* and *Hemigalus* than it is to such relatively primitive viverrids as *Fossa*. Each of the lower teeth p_3, p_4, m_1 of *Chrotogale* appears to be merely an anteroposteriorly elongate but enfeebled derivative of the corresponding tooth in *Diplogale* or *Hemigalus*. In m_1 the strong large blades of the carnassial of primitive viverrids are here represented by the slender pricking cusps of the trigonid, while the short talonid of primitive viverrid carnassials contrasts with the compressed talonid in this specialized member of the Hemigalinae.

The deciduous dentition of *Chrotogale*, as figured by Oldfield Thomas (1912, pp. 501, 502), foreshadows the adult dentition instead of retaining primitive viverrid characters. This is true of both the deciduous incisors and of the cheek teeth.

The ectotympanic has an enlarged oval orifice and a much more prominent ventral rim than in *Hemigalus*. The occipital condyles are remarkably wide (see Table 7). The median occipital swelling for the vermis cerebelli is more prominent than in *Hemigalus*.

According to Pocock (1933, p. 1000), "Evidence for the perfume-glands in *Diplogale* and *Chrotogale*, similar to those described in the ♂*Hemigalus* by myself and in ♀ by Mivart, is to be found in some made-up skins of those genera."

Fossa in nearly all characters is a good structural link between the primitive Viverrinae and the Hemigalinae. In *Cynogale*, one of the specialized derivatives of *Hemigalus*, there has been a marked narrowing across the interlacrymal diameter, the forehead and the palate, while the olfactory chamber has become longer (Table 6).

TABLE 6

FOSSA, HEMIGALUS, CYNOGALE

Skull Proportions	Index	Assumed Primitive *Genetta*	Range			Remarks
			Fossa	*Hemigalus*	*Cynogale*	
Total skull length		82	93	92	118	Considerable increase
Bizygomatic width	I	50.0	53.4	51.1	55.1	Nearly same
Max. cran. width	VI	54.7	60.7	60.4	60.9	Nearly constant
Forehead width	VIII	41.4	38.1	40.6	25.6	Becoming very narrow
Interlacr. width	II	61.5	58.2	52.8	44.2	Becoming very narrow
Palatal width	X	73.0	60.2	60.9	55.0	Becoming very narrow
Lower dental arch width	XIV	54.1	48.0	45.7	49.2	Apparent narrowing due to increased jaw length
Bicondylar width			39.4	39.4	41.4	Condyles becoming very wide
Olfactory length	III	54.7	76.4	73.6	84.4	Olf. chamber becoming long
Orbital length	IV	34.0	35.6	30.1	29.7	Orbits slightly reduced
Basicranial length	XI	47.2	42.1	43.4	43.8	Remarkably constant
Upper carnassial width		70.5	78.8	101.6	91.7	Progressively procyonoid
Upper trigonid length m_1		68.4	61.1	47.9	45.3	

Cynogalinae

Another aberrant offshoot of the hemigaline stock is the "otter civet" of Borneo, *Cynogale bennettii*, which is made the type of a subfamily (Cynogalinae) by Pocock. The entire skeleton of this semi-aquatic viverrid has been well figured in De Blainville's "Osteographie," and we owe to the School of Dental and Oral Surgery of Columbia University the opportunity of studying a well preserved skull. The skeleton is remarkable for the strength of the fore limbs and the reduction of the tail. In the general morphology of its dentition *Cynogale* is closely allied with *Hemigalus* but with an increase in size and emphasis of the accessory cusps on the slopes of the compressed upper and lower premolars, producing serrate crowns which are not wholly unlike those of the fish-eating phocids. The upper cheek teeth ($p^4 - m^2$) have lost or reduced their trenchant blades and have acquired rounded lingual ridges that bear several cuspules. The deciduous cheek teeth, as figured by Leche (1915, p. 303), are small, compressed and cuspidate. The fourth lower premolar is secondarily compressed and serrate and has probably lost the sub-carnassial trigonid which this tooth bears in *Hemigalus*.

The skull of *Cynogale* is much more robust than that of *Hemigalus*, with heavier arches, larger muscle crests, much sharper postorbital constriction, secondarily narrowed forehead, entotympanic extending

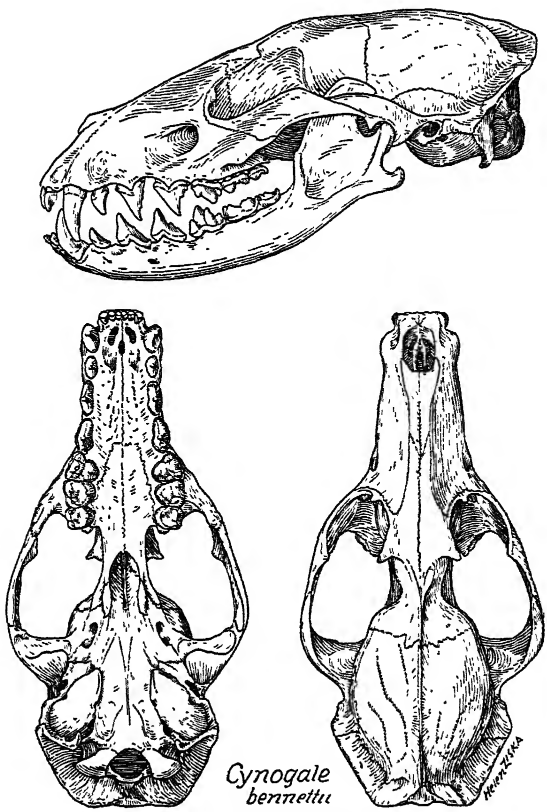


FIG. 21. *Cynogale*, three views of skull. × ½.

farther forward, paroccipital process extending below level of bulla; the infra-orbital foramen is much enlarged, in correlation with the abundant development of facial vibrissae, while the areas of origin of the muscles of the snout in front of the anterior border of the orbit are emphasized. All this, together with the valve-like form and dorsal position of the nostrils, is part of the amphibious, predatory habitus.

After an ingenious analysis of the evidence from the characters of the facial vibrissae, feet, teeth, etc., of this animal, Pocock (1915*a*, pp. 359, 360) infers that it probably lies submerged with only the tip of the nose exposed above the surface, frightening the fish into crevices and holes, thrusting its sensitive snout in after them, cutting them up with its serrate lower premolars and chewing them with its blunt molars.

Its feet, according to the same authority (1915*a*, p. 357), are nearest to those of the arboreal Paradoxurinae, the manus being relatively wider than that of *Hemigalus*. Its perineal scent glands are somewhat reduced.

In conclusion, the subfamily Cynogalinae may be defined by the amphibious and piscivorous habitus characters noted above, including: (*a*) the incipient widening of the feet, (*b*) the shortness of the tail, (*c*) the dorsal position and valvate form of the nostrils, (*d*) the multiplication of the facial vibrissae, (*e*) the piscivorous characters of the jaws and dentition. But these habitus characters merely disguise and do not obliterate the many heritage characters which *Cynogale* has retained from the hemigaline stem. If taxonomists chose to stress phylogenetic relationships rather than differences, the subfamily Cynogalinae would assuredly be merged in the Hemigalinae.

Euplerinae

The rare *Eupleres goudoti* of Madagascar is made the type of a subfamily, Euplerinae, by Pocock, but a careful comparative study indicates that it is only an aberrant side branch of the Hemigalinae. The external appearance of *Eupleres* in the color plate published by Gray (1870, Pl. LI) shows a considerable general agreement with that of the North Bornean *Diplogale hosei* as figured by Oldfield Thomas (1892, Pl. XVIII). In both the general body color is fairly uniform, lacking the sharp blotches and streaks that are characteristic of many Viverridae. Both have small heads with pointed muzzles, large necks, long bodies and fairly long tails. But there are many differences in the details of color and proportions, *Eupleres* being more specialized in its olive-colored fur, longer muzzle, and great bushy tail, which is somewhat suggestive of *Myrmecophaga*-like habits.

Oldfield Thomas (1912, pp. 499, 502) in describing the juvenile skull, jaws and teeth of *Chrotogale owstoni* noted several resemblances to those of *Eupleres* but evidently did not regard them as indicative of close relationship. We, however, cannot set aside as insignificant the striking agreement in fundamental plan, disguised as it is by differences in detail. Although our skull of *Eupleres goudoti* is definitely that of a younger animal than that figured by Gray (1870, p. 827), the resemblance to the immature skull of *Chrotogale owstoni* figured by Oldfield Thomas (1912, pp. 500–502) is very striking. In both genera the braincase is well rounded, elongate, forehead relatively wide, zygomata weak, sagittal and lambdoidal crests absent, muzzle tapering in front, tip of muzzle depressed below plane of cheek teeth, external auditory meatus

forming a backwardly directed oval; palate narrow, cheek teeth feeble, $p^4 - m^1$ triangular, anterior palatal and median septal foramina forming a triad of long oval slits; basis cranii long, occipital condyles very wide, median vertical convexity for the vermis cerebelli very conspicuous; mandible elongate, slender, with procumbent incisors and compressed cheek teeth.

According to Beddard (*Cambridge Natural History*, p. 403), the animal "appears to burrow in the ground and possibly contents itself with a diet of worms" but grubs and grasshoppers have also been suggested as possible alternates.

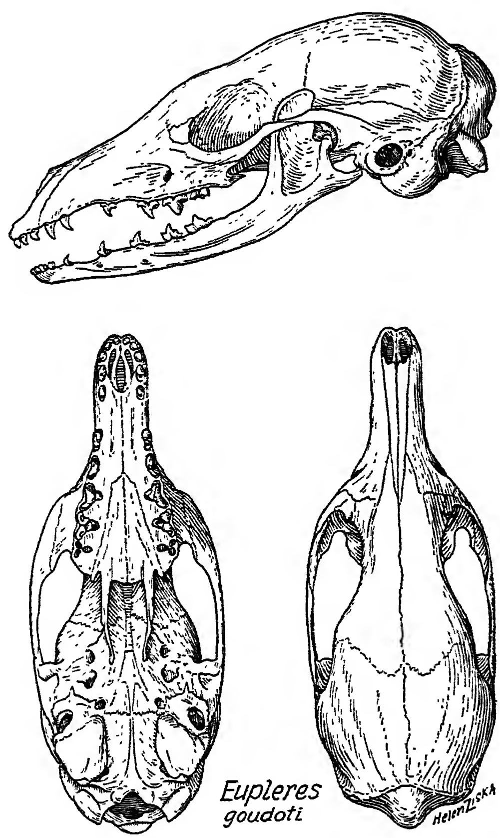


FIG. 22. *Eupleres*, three views of skull. × ¾.

As figured by J. E. Gray (*P. Z. S.*, 1870, p. 824, Pl. LI), *Eupleres* has advanced beyond *Chrotogale* by emphasizing and distorting certain features of the immature skull and dentition of that genus: (1) the slenderization and elongation of the muzzle, partiuclarly in the region

between p^1 and p^2, which are separated by a much widened diastema; (2) the narrowing of the front part of the muzzle; (3) the secondary convergence (polyisomerism) in form of the incisors, canines and p^1, which are all pointed, recurved hooks; (4) the spatulate arrangement of the lower incisors and canines, the latter having been taken over into the incisor series as in ruminants; (5) the change in position of p_1, which has moved forward near the canine, widening the diastema between it and p_2, whereas in *Chrotogale* p_1 is nearer to p_2 than to the canine; (6) with the spatulate development of the lower front teeth the tip of the greatly diminished lower canine is withdrawn lingual to the line connecting i^3 and the upper canine; (7) the upper cheek teeth $p^4 - m^2$ of *Eupleres* differ markedly in appearance from those of *Chrotogale*, being far more slender, with deeply notched buccal face, very slender internal cusp (protocone) and restricted central fossa on m^1.

If we knew only the p^4, m^1, m^2 of *Eupleres* and of *Chrotogale*, we would not suspect that in many other characters of the dentition and skull the former appears to be derived almost directly from the latter; nevertheless in view of this prevailing morphological agreement between the two genera, it is more than conceivable that the cheek teeth of *Eupleres* have been derived from those of *Chrotogale* or some near relative of the latter by the following changes: (*a*) rapid decrease in size of the crowns and roots as a whole; (*b*) rapid reduction of the external cingula and hypocones in the upper, and of the talonid in the lower, molars; (*c*) the tendency for the premolars to be spaced and for the inner parts of the upper molars and the posterior parts of the lower molars to be rapidly reduced, thus widening the interdental embrasures in the upper to provide space for the relatively enlarged trigonids of the lower; (*d*) with the relative enlargement of the trigonids, the parastyle and metastyle of p^4, m^1 have become more pronounced, producing what we may call a false carnassial rather than a primitive carnassial form. As noted by Leche, rather similar differences distinguish such specialized long-snouted genera as *Tupaia* and *Hemicentetes* from their respective allies.

The deciduous cheek teeth of *Eupleres*, as figured by Leche (*Zool. Jahrb.*, XXXVIII, p. 290) resemble their successors of the permanent set more closely than is usually the case. Nevertheless they bear a striking general likeness to those of the immature *Chrotogale* as figured by Oldfield Thomas (1912, p. 501). Even the deciduous incisors, canines and $p\frac{1}{1}$ of *Eupleres* may be conceived as degenerate derivatives of the more complete dentition of the immature *Chrotogale*. In other words, the dentition as well as the skull of *Eupleres* may have been de-

rived in part from that of *Chrotogale* by a process of infantilization, or arrest of development, accompanied by a marked change in general trend of development.

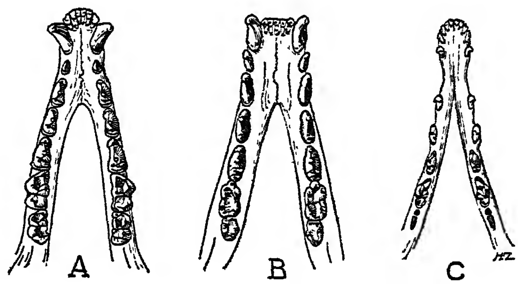


FIG. 23. Lower jaws of Hemigalida.

A. Fossa. × ⅔.
B. Cynogale. × ½.
C. Eupleres goudoti. × ¾.

In short, all that is needed to transform the *Chrotogale* into the *Eupleres* skull and dentition is a change in relative growth rates, some being delayed, others hastened (anisomerism).

Inspection of Pocock's figures (1915e, Pl. XIII) of the feet of *Eupleres* shows that all five digits are retained on both manus and pes and that the pollex and hallux are undiminished. Pocock (1915e, p. 349) regards the comparatively large size and low position of the hallux and pollex as a primitive feature, but we would interpret the evidence as consistent with our conclusion that in respect to its feet *Eupleres* is a somewhat badger-like or semi-fossorial derivative of the primitive hemigaline stock, the manus having been secondarily shortened and broadened and the first digit in both manus and pes moved forward and secondarily enlarged, as have also the claws of all the digits. Such a conclusion will doubtless appear unwarranted to those who may be unacquainted with the evidence that even the "plantigrade" feet of badgers and bears are less primitive than the digitigrade feet of the ancestral viverrid-canid stock. Moreover, both the manus and pes of *Eupleres* appear to us to be readily derivable from those of *Hemigalus derbyanus* (Pocock, 1915c, Pl. VIII), the latter in turn being derivable from the central semi-digitigrade feet of the Viverrinae.

In her analysis of anatomical observations on *Eupleres*, Carlsson (1902) lists nine characters which this form shares with the Viverrinae, eight in common with the Herpestinae and fourteen peculiar to itself. She did not, however, compare it especially with *Hemigalus*. According to our observations, the dentition of *Eupleres* is widely different from

that of any of the Herpestinae and the same is true of its mandible and skull. The dentition of *Eupleres* appears to be readily derivable by easily understandable steps from that of *Chrotogale* by the reduction in size of the cheek teeth, by their wide spacing and by their anteroposterior lengthening. The bulla may be regarded as an infantilized derivative of that of *Chrotogale*, in which the ectotympanic has remained more ring-like; it differs from those of the Herpestinae in having a sharp groove between the ecto- and entotympanic. The occipital region is quite near to that of primitive Viverrinae, particularly in its median convexity for the vermis cerebelli. But *Eupleres* is much nearer to *Hemigalus* and *Chrotogale* in skull and dentition, taken as a whole, than it is to *Genetta* or any other of the Viverrinae.

TABLE 7

FOSSA, HEMIGALUS, CHROTOGALE, EUPLERES

Skull Proportions	Index	Assumed Primitive *Genetta*	Range				Remarks
			Fossa	*Hemigalus*	*Chrotogale*	*Eupleres*	
Total skull length..		82.0	93.0	92.0	107.0†	80.0*	Increase in length followed by reduction
Bizygomatic width.	I	50.0	53.4†	51.1	48.0	42.5*	Decrease with reduction of jaw muscles
Max. cran. width..	VI	54.7	60.7	60.4	60.3*	66.7†	Slight increase correlated with reduction of jaw muscles
Forehead width....	VIII	41.4	38.1*	40.6	43.6	53.1†	Widening, with reduction of jaw muscles
Interlacr. width...	II	61.5	58.2	52.8	51.7	51.4	Subfamily remarkably constant
Palatal width.....	X	73.0	60.2	60.9†	51.8	46.2*	Marked decrease correlated with weakening of jaws and teeth
Lower dental arch width	XIV	54.1	48.0†	45.7	44.1	35.1*	Marked decrease correlated with weakening of jaws and teeth
Bicondylar width..	XVI		39.4	39.4	42.1	47.0	
Olfactory length...	III	54.7	76.4	73.6	77.4†	66.7*	Relative decrease in *Eupleres*
Orbital length.. ..	IV	34.0	35.6	30.1	30.1	34.2	Fairly constant
Basicranial length..	XI	47.2	42.1	43.4	40.1	45.8	Fairly constant
Upper carnassial width..........		70.5	78.8	101.6	80.6	62.8	
Lower carnassial length..........		68.4	67.1	47.9	47.1	70.6	

The figures show by what relatively small steps one may pass from *Fossa* to *Hemigalus* to *Chrotogale* and finally to *Eupleres*. In the latter there has been a marked decrease in bizygomatic width, a widening of the forehead, and a marked decrease in the width of the palate and lower dental arch.

* Lowest, † highest.

The foregoing tables (6, 7) are entirely consistent with the following conclusions:

(*a*) That *Fossa* stands near the base of the hemigaline series;
(*b*) That *Cynogale* is closely related to *Hemigalus*, differing by long muzzle, very narrow forehead, narrow interlacrymal width, narrow palate, wide condyles;
(*c*) That *Chrotogale* is close to *Hemigalus* but with long muzzle, narrow palate, wide condyles;
(*d*) That *Eupleres* is close to *Chrotogale*, but with low bizygomatic width, very narrow palate, wide cranium and extremely wide condyles.

The Ancestry of the Galidictines

Family Viverridae (*continued*)

Section *Galidictida*

Subcarnivorous derivatives of Viverrida, superficially approaching the broad-skulled herpestids in skull form; cheek teeth less carnassial than in stem Viverrida, with relatively wider upper carnassial and larger talonid on m_1. *Entire skull relatively wide and low; bony muzzle short, wide and low; forehead (supraorbital region and postorbital constriction) wide; ectotympanic somewhat inflated, with prominent spout; entotympanic inflated transversely, flat ventrally. Feet derived from the Genetta type but with non-retractile claws and pollex near to digit II, its pad near to the trilobed pads of digits II–V. Perineal pocket and glands much as in Paradoxurus (but present only in the female); prepuce far in advance of scrotum (as in Fossa and Cryptoprocta).*

Galidictinae

The Galidictinae are in many respects structurally intermediate between the Viverridae (in Pocock's sense) on the one hand, and his Mungotidae (our Herpestidae) on the other. Carlsson (1910) after an extended analysis of the anatomy of these little animals, which, like so many other "living fossils," are native to Madagascar, lists seven characters which one or another of them shares with the Viverrinae, ten which they have in common with the mongoose subfamily (Herpestinae) and fourteen others which are peculiar to this subfamily. In respect to the general appearance of the skull and dentition these forms have a strong claim to cousinship with the mongoose group, particularly in the construction of the auditory bulla. On the other hand, because they possess perineal scent glands like the typical Viverridae and do not stress

the anal glands as do the Herpestidae, Pocock separates them from the latter and refers them to a distinct subfamily.

The skull of *Galidictis* has a large, wide braincase, short muzzle and very short, powerful jaws, the whole effect being not unlike that seen in the skunks and otters. The lower canines are remarkably robust for the size of the jaw and the little animal must be able to hold fast to a

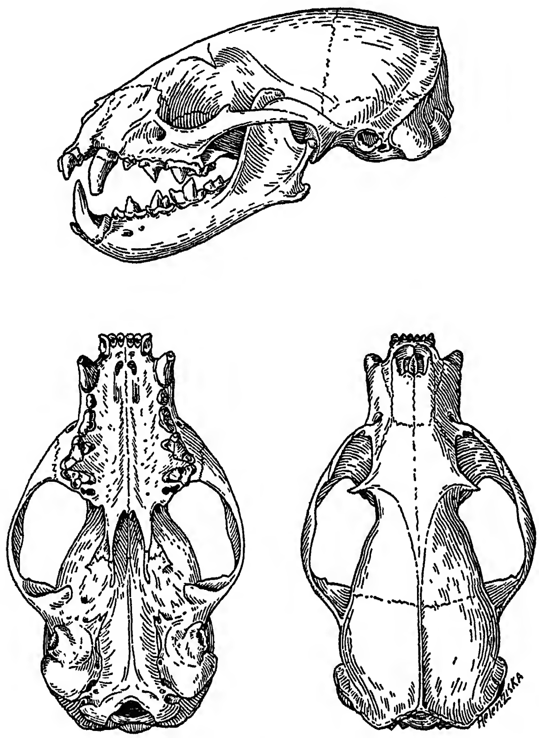


FIG. 24. *Galidictis*, three views of skull. × ¾.

relatively large struggling prey, such as a snake. The upper carnassial has a well developed blade but the crown as a whole is wider transversely and shorter anteroposteriorly than in the primitive Stenoplesictinae. The first upper molar is small and the second much reduced and narrow anteroposteriorly. The cheek teeth are not far from those of *Herpestes*, which likewise inherits much of the stem carnassial stamp.

Although skulls of *Galidia* strongly suggest those of herpestids (*e.g.*, the almost backwardly pointing orbital notch in *Galidia*, as in *Mungos*, the ectotympanic spout, as in herpestids), yet no one herpestid genus seems to be close to any Galidictinae and it may well be that Pocock is right in separating the Galidictinae from the Herpestidae and treating it as a subfamily of Viverridae.

The color pattern of the fur (Pocock, *Ann. Mag. Nat. Hist.*, Aug., 1915, Pl. VII; Oct., 1915, p. 356) is much closer to that of the genets, making no approach to the ticked fur and transverse bars of herpestids (in this it may be only more primitive). As to the feet, the pollex is nearer to digit II and its pad near to the trilobed pad of digits II–V, as in the genets, whereas in herpestids if the pollex is present its pad is well separated from the trilobed pad. Similarly in the pes. According to Pocock (1915*f*, p. 354, 355) also the claws are non-retractile (as in the mongooses). The Galidictinae have well developed perineal pocket and glands much as in *Paradoxurus*, while the anal pouch is absent (unlike the Herpestidae). Also the prepuce is far in advance of the scrotum, as in *Fossa* and *Cryptoprocta* and hence wholly unlike the herpestids. On the whole, the Galidictinae may prove to be derivatives from the stem of the Paradoxurinae, not far from *Fossa*.

Comparison of the skull of *Galidictis* with those of *Galidia, Fossa, Arctogalidia*, brings out the following points: (1) The Galidictinae may be pygmy derivatives of the stem of *Fossa*, with secondarily shortened and widened muzzles, greatly widened foreheads, shortened tooth rows and widened palate. (2) As in small skulls generally, the bullae are

TABLE 8

GALIDICTINAE COMPARED WITH *Genetta* AND *Poiana*

Skull Proportions	Index	Assumed Primitive *Genetta*	*Poiana*	*Galidia*	*Galidictis*	Remarks
Total skull length		82.0	61.0	58.0	62.5	Skulls ¼ shorter than in *Genetta*
Bizygomatic width	I	50.0	55.7	60.3	65.6	Becoming wider
Max. cran. width	VI	54.7	62.5	60.5	64.9	Becoming wider
Forehead width	VIII	41.4	56.0	61.5	49.6	Becoming wider
Interlacr. width	II	61.5	70.0	85.7	95.0	Becoming very wide
Palatal width	X	73.0	70.4	80.8	91.5	Becoming very wide
Lower dental arch width	XIV	54.1	53.8	59.3	64.9	Becoming very wide
Lower jaw spread	XII					
Bicondylar width						
Olfactory length	III	54.7	52.5	45.0	38.9	Becoming very short
Orbital length	IV	34.0	35.0	27.5	27.3	Somewhat reduced
Basicranial length	XI	47.2	47.5	47.5	43.6	Nearly constant
Upper carnassial width		70.5	55.7	81.5	75.0	Becoming wider
Lower carnassial length		68.4	80.0	60.7	64.0	Talonid increasing

expanded but the spout-bearing ectotympanic is fundamentally the same as in *Fossa*. A juvenile *Paguma* (subfamily Paradoxurinae) has an ectotympanic spout nearly as well developed as in the Galidictinae. (3) The elongate cranial vault of *Galidia* is not unlike the juvenile

Paguma. (4) The proximally broad nasals of the Galidictinae are more advanced than those of *Paguma* but differ widely from the narrow V-shaped nasals of the Herpestidae. (5) The horizontal dorsal stripes in *Mongotictis* (of the subfamily Galidictinae) are comparable with the dorsal stripes of the Javan palm civet (*Arctogale trivirgata*) (see Lydekker, *Natural History,* Vol. I, Sect. 2, p. 453).

We therefore place the Galidictinae as an offshoot from the base of the viverrid stem where it joins the herpestid branch and not far from *Fossa* and the line leading to the Paradoxurida.

The two genera of Galidictinae represented go far beyond *Poiana* (partly parallel form in the Viverrinae) in increasing the transverse diameters of the forehead, cranial, interlacrymal, palatal and lower dental arch; their olfactory chamber becomes very short but the "basi-cranial" length (in proportion to brain length) remains nearly constant. The relative carnassial width has increased and so has the talonid length.

TABLE 9

Galidictis Compared with *Genetta*, *Fossa* and the Smaller Herpestines

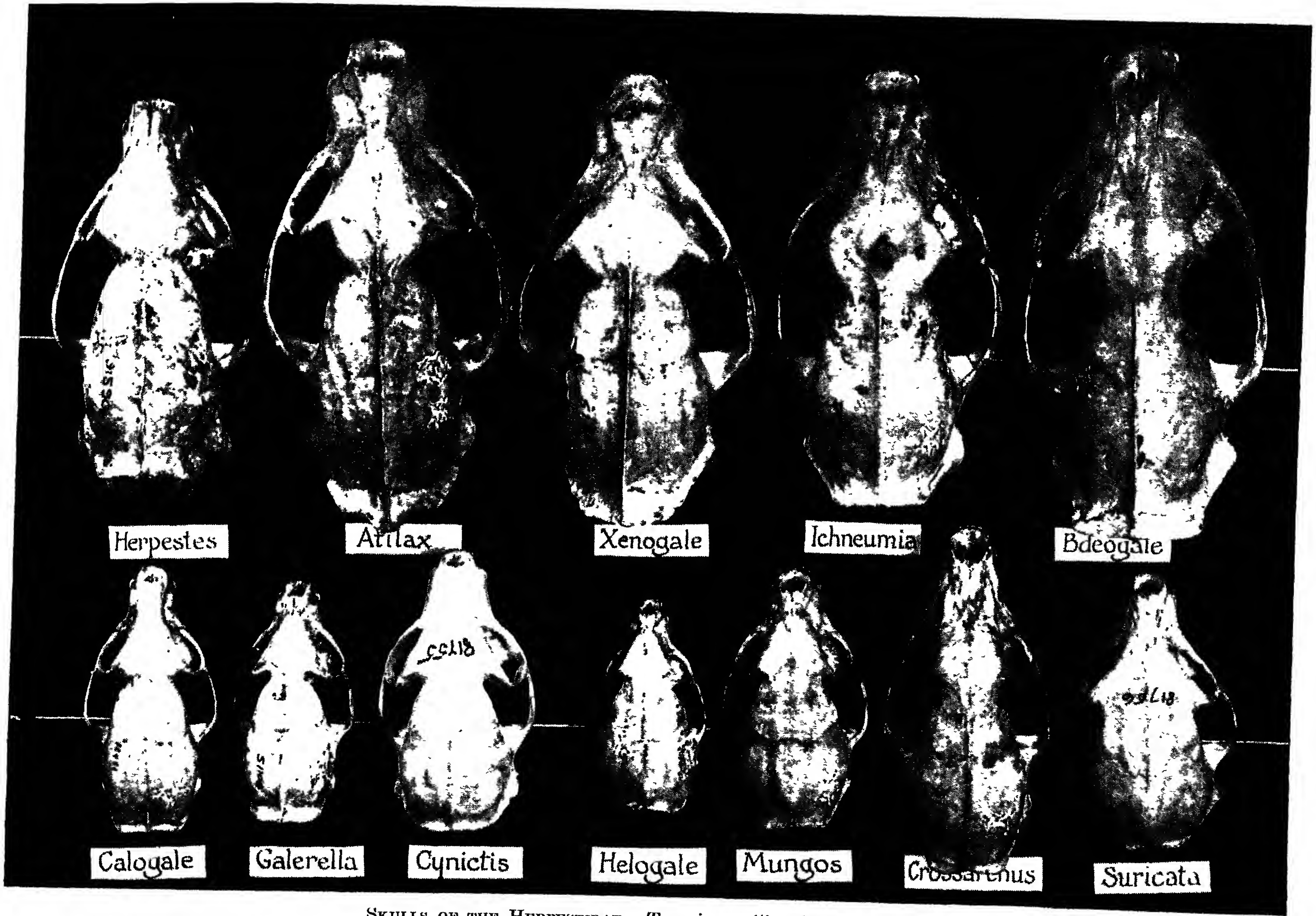
Skull Proportions	Index	*Genetta*	*Fossa*	*Galidictis*	*Calogale*	*Cynictis*	*Suricata*
Total skull length		82.0	93.0	62.5	60.9	59.0	55.0
Bizygomatic width	I	50.0	53.4	65.6	52.9	62.7	80.0
Max. cran. width	VI	54.7	60.7	64.9	58.3	67.5	78.9
Forehead width	VIII	41.4	38.1	49.6	48.1	59.2	70.0
Interlacr. width	II	61.5	58.2	95.0	72.2	85.7	72.7
Palatal width	X	73.0	60.2	91.5	77.2	81.5	95.8
Lower dent. arch width	XIV	54.1	48.0	64.9	55.4	58.6	76.0
Bicondylar width							
Olfactory length	III	54.7	76.4	38.9	48.5	47.5	44.7
Orbital length	IV	34.0	35.6	27.3	24.9	32.5	34.2
Basicranial length	XI	47.2	42.1	43.6	47.8	45.0	34.2
Upper carnassial width		70.5	78.8	75.0	77.4	86.9	120.0–138.7
Lower carnassial length		68.4	67.1	64.0	61.8	55.2	51.1– 66.7

Although the *Galidictis* skull is nearer to the small herpestine skulls than to the "primitives" in respect to total length and to the indices of the bizygomatic, max. cranial, forehead width and olfactory length, yet it differs from the herpestines in other important features not shown in the table (see text), so that its metrical resemblances to the latter, which are rather scattering, may be partly due to parallelism and in any case are not close enough to indicate near relationship.

The Ancestry of the Mongoose Family

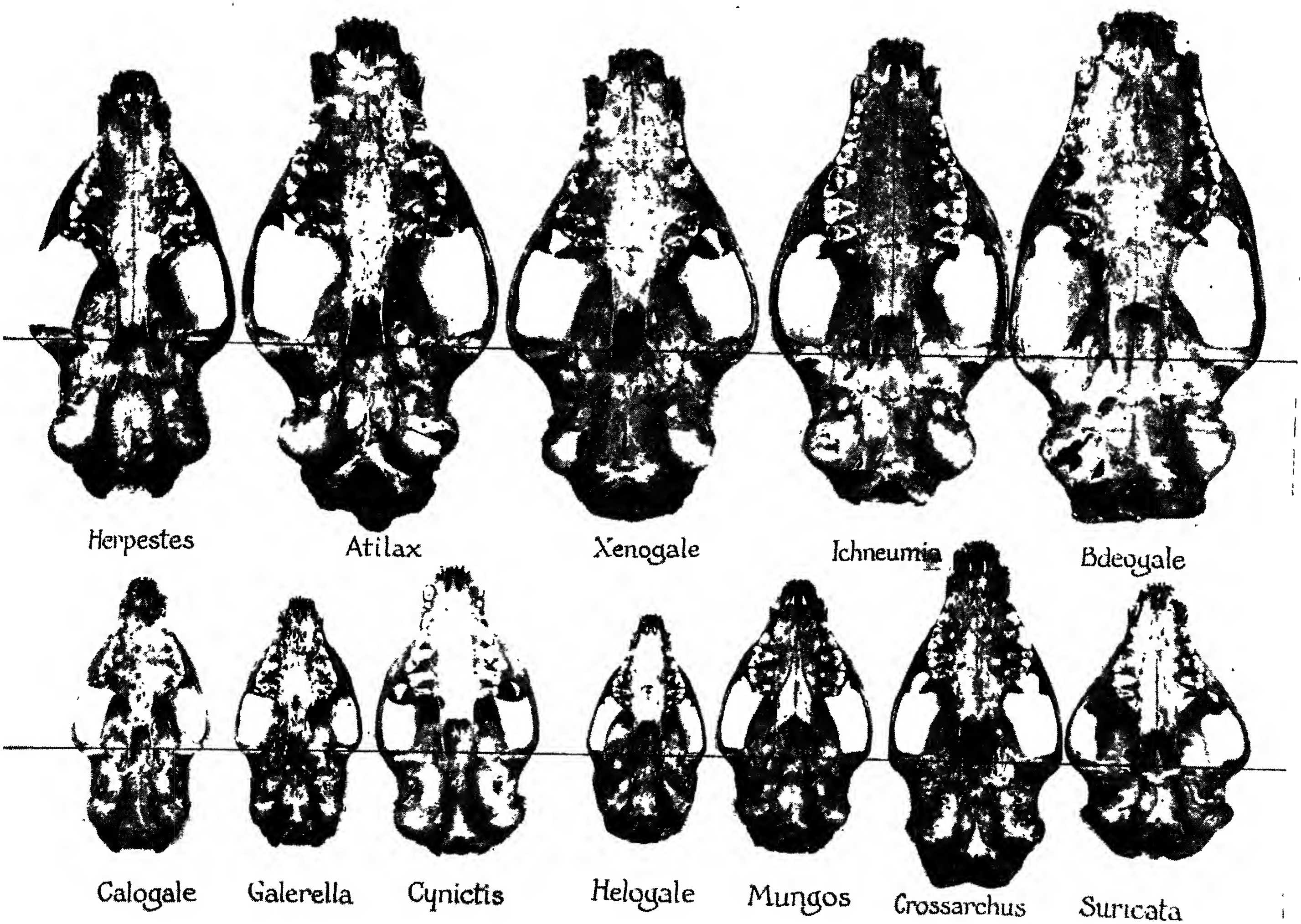
Family Herpestidae

The mongooses (family Herpestidae) are small, mostly short-legged mammals, prevailingly terrestrial or semi-fossorial, with long non-



SKULLS OF THE HERPESTIDAE. Top view. *Circa* ¾ natural size

PLATE VI



SKULLS OF THE HERPESTIDAE. Underside view.

retractile claws, narrow feet, often with pollex and hallux reduced or absent. They have no perineal glands but the anal glands are well developed and open into a deep pouch. In *Herpestes* and *Calogale* the braincase is very long and narrow, the muzzle short and high, the relatively small orbits completely ringed posteriorly. The ectotympanic has a more or less spout-like external auditory meatus, finally with a transverse slot or depression on its lower surface. The cheek teeth and jaws of the more central genera are of the general carnassial type noted above under *Galidictis* but with less enlarged canines. Such is the dental apparatus of these famous snake-killers, eminently fitted to inflict small but grievous bites in the heads of serpents.

In *Bdeogale,* a large African relative of the mongoose, the posterior cheek teeth p^4, m^1, have largely lost their trenchant characters and now bear low tubercles and blunt blades somewhat recalling the conditions in *Hemigalus* and *Arctictis*. The African explorer Peters is cited as having found the remains of a rhinoceros viper in the stomach of a *Bdeogale,* so that although the cheek teeth do not look much like those of *Herpestes,* both animals can apparently kill poisonous snakes. And though the *Bdeogale* skull is much larger than that of *Herpestes,* the relationship between the two is indicated by the essentially common construction of the compound auditory bulla, convex frontals, short muzzle, etc.

The most specialized skull form among the herpestines is that of the mierkaat, *Suricata,* in which the braincase is extremely wide, the bullae inflated, especially at the posterior end; the jaws are very short and the cheek teeth (p^4, m^1) much shortened anteroposteriorly and widened transversely, so that they approach in appearance those of the extinct insectivore *Ictops*. Insects are indeed said to form an important item in their daily food.

Within the group there are about a dozen genera which collectively afford an astonishingly wide range of variation in the proportions of the skull (Tables 9, 10) and in the adaptive characters of the cheek teeth, while maintaining a singular degree of unity in their subfamily heritage.

The herpestid genera may be grouped according to the adaptive trends of their dentition and skulls as follows:

A. Herpestes, Atilax, Xenogale, Ichneumia, Bdeogale.

Size large; cheek teeth from carnassial to procyonine, with progressively enlarging m^1 and m^2; p^4, m^1 with diminishing shear blade and decreasing inner carnassial angle; para- and metaconid of m_1 originally large, blade-like and widely separated, becoming reduced and connate; p^1_1 persistent; palate widening in front, endocranium elongate; ectotym-

panic less inflated ventrally than entotympanic; orbits directed chiefly outward.

B. Calogale, Galerella, Cynictis.

Size small; persistently carnassial; $p\frac{1}{1}$ persistent but small; m^2 much reduced; endocranium widening, postorbital processes directed sharply backward and downward; ectotympanic inflation extending ventrally below level of entotympanic inflation; palate widening posteriorly.

C. Helogale, Crossarchus, Mungos, Suricata.

Size small; cheek teeth ($p^4 - m^2$) becoming ictopsine (*i.e.*, antero-posteriorly short, transversely wide), p^4 with diminishing metastyle blade and increasing protocone, m^1 becoming wider and more symmetrical, outer carnassial angle opening up, inner carnassial angle narrowing; m^2 becoming like m^1 but remaining smaller; paraconid and metaconid of m_1, originally large and blade-like, becoming reduced and connate; $p\frac{1}{1}$ absent; skull becoming very wide and short; palate widening posteriorly; ectotympanic becoming depressed and widened, finally with wide transverse groove or slot.

Group A

Herpestes, with many species, appears to be the central genus for the entire family. Its very narrow skull is advanced in its vertically deep but short olfactory chamber, well developed orbits, long capacious braincase and expanded middle ear, combined with short powerful jaws equipped with highly effective teeth for killing and shearing the prey. The forehead forms a convex protuberance, due to the dorsad inflation of the naso-turbinates. The muzzle is short but high; the well-rimmed orbits look forward and outward. The postorbital constriction is well marked, the cranial vault long and narrow. The entotympanic chamber is moderately inflated transversely and the ectotympanic bears a foramen on its ventral surface and is produced laterally into an incipient spout. The ectotympanic is considerably less extended and inflated ventrally than the entotympanic.

The cheek teeth are of the sectorial type, $p\frac{1}{1}$ persistent, p^4 having a prominent metastyle shear which meets the anteriorly produced parastyle of m^1 at a marked angle. M^1 is extended transversely with sharp cusps, m^2 is almost vestigial; m_1 has a carnassial trigonid with large secant proto- and paraconids; the metaconid is retained, the talonid relatively small and narrow; m_2 is very small with a high small trigonid and basin-shaped talonid. This dentition is but little removed from that which seems to be the central type among the upper Eocene viver-

rids; it differs from that of a putative ancestor of the Canidae, *Cynodictis compressidens,* in that the upper molars have no internal cingulum. This dentition is also close to that of the fossil *Viverra simplicidens* but

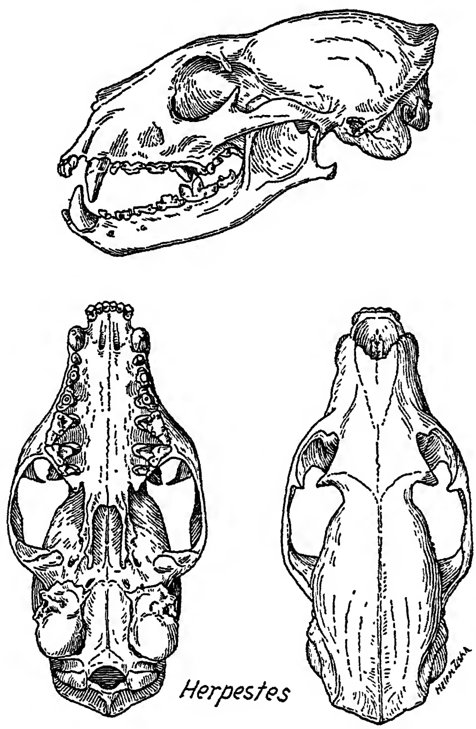


FIG. 25. *Herpestes,* three views of skull. × ½.

is somewhat more specialized. Digits 5–5, the hind foot long with naked heel; tail long and tapering.

Atilax macrodon, the marsh mongoose of the Belgian Congo. Size very large. Cheek teeth large and massive, the upper p^4, m^1, with conic, procyonoid cusps; p^4 with smaller metastyle blade and massive inner cusp; m^1 large, with bucally prominent parastyle, conic metacone and expanded protocone; p^1 absent; p^2 enlarged; upper canine straight, stout, dagger-like. The cheek teeth are somewhat hyaena-like in their massiveness. Lower jaw powerful, with massive recurved canine and expanded rounded premolars; p_3 with prominent bucally-projecting hypoconid; m_1 with paraconid moved posteriorly, large wide talonid; $p\frac{1}{1}$ absent. Entotympanic much inflated, globose. Digits 5–5, exceptionally long and separate. Heel long, not hairy. Pollex and hallux

just above level of interdigital pads. Tail moderate. Habits partly otter-like. A derivative of *Herpestes* closely related to *Xenogale*.

Xenogale microdon. The animal itself, as figured by Lang (Allen, 1924, Pl. 27) is similar to *Herpestes* but the nose is more projecting. The

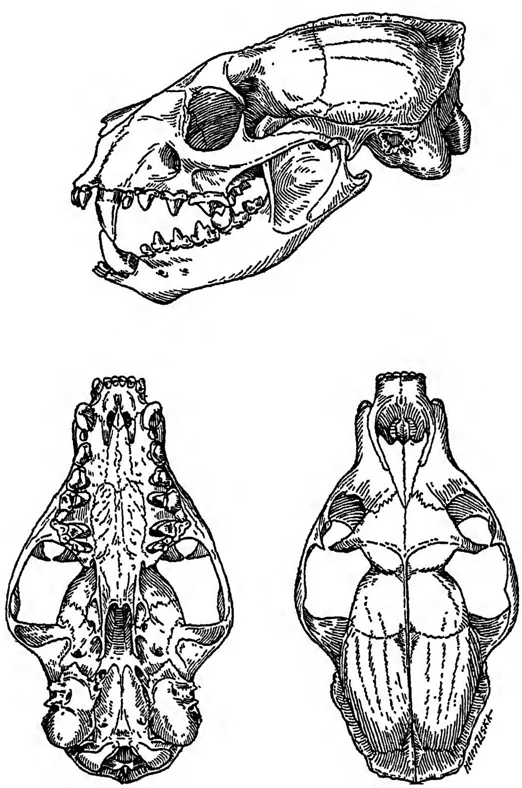


FIG. 26. *Xenogale*, three views of skull. × ½.

skull is somewhat larger than that of *Herpestes* but with the forehead less inflated and the muzzle larger. The prominent upper canine is slightly curved and dagger-like, with a marked postero-external vertical groove. The palate is wider than that of *Herpestes*, especially in front. P^4 and m^1 are still of carnassial type but m^1 is longer, as is also m_2; p_1^1 is minute. Digits 5–5. Hand wide, heel hairy. Tail thick and of moderate length. A derivative of *Herpestes*, but showing the initial stage in the transformation of the cheek teeth from the carnassial into the procyonoid or crushing type. No. 1515, D. C. A., Amer. Mus. Nat. Hist., *Xenogale* sp., tends to connect this genus with *Herpestes*.

Ichneumia. A large conic-toothed derivative of *Herpestes*, parallel to but less advanced than *Bdeogale*, retaining the elevated inflated fore-

head but widening the palate, p^1_1 present, enlarging the internal cusp of p^4, reducing its metastyle, enlarging m^1, m^2, making m^1 more symmetrical with less laterally projecting parastyle; anterior root of zygomatic arch far behind p^4; m_1 with connate paraconid and metaconid; m_2 elongate with six cusps. Upper canine slightly recurved, less dagger-like; entotympanic much inflated ventrally. Digits 5–5, pollex and hallux small, well above level of interdigital pads, heel hairy. Tail very large and bushy.

Bdeogale. A procyonoid or blunt-toothed derivative of *Xenogale*, closely related also to *Ichneumia*. P^1_1 present. Palate enlarged, widened interiorly, muzzle massive. Cheek teeth procyonoid. Anterior root of zygomatic arch more posterior than in *Herpestes*, it being behind the contact of p^4 and m^1 and opposite the contact of m^1 and m^2. P^4 with

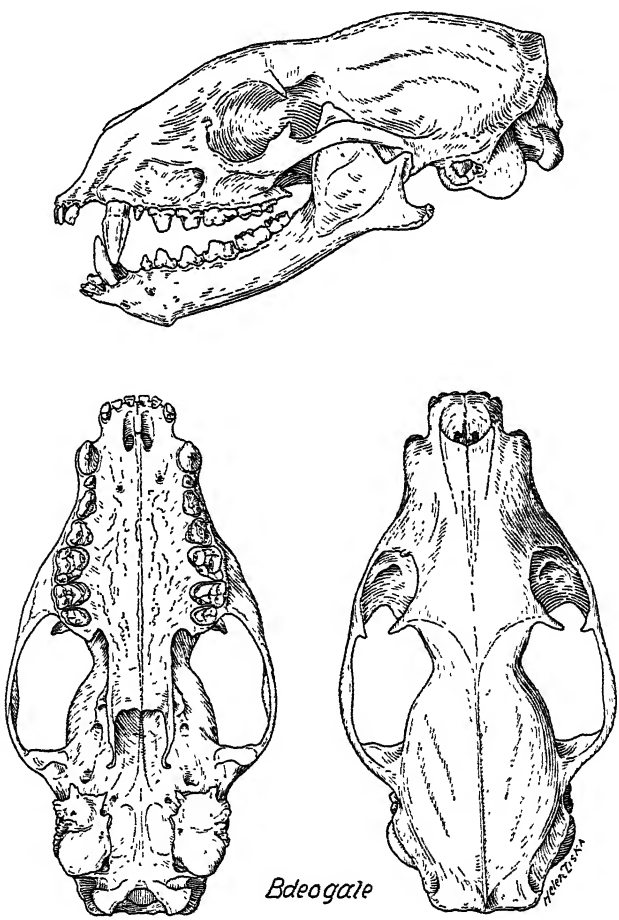


FIG. 27. *Bdeogale*, three views of skull. × ⅔.

much reduced conical metastyle shear, prominent parastyle, enlarged conical internal cusp and distinct accessory cusp (tetartocone) on postero-internal cutting edge; thus the outer part of the triangular interdental embrasure is squeezed out by the submolariform shape of p^4; m^1 large, with bluntly conical paracone and metacone, the paracone more in line with the metacone; swollen protocone; outer carnassial angle between p^4 and m^1 obtuse; m^2 relatively much larger than in primitive types, with swollen protocone; the inner moiety of the crowns of p^4, m^1, m^2 has increased at the expense of the outer moiety; m_1 without carnassial blade, the para- and metaconid connate or twinned, the hypoconid conic; m_2 unusually large and functional, tending to be like m_1. Feet wide, derived from *Xenogale* type by loss of pollex and hallux; digits 4–4; heel hairy. Tail bushy.

TABLE 10

THE LARGER HERPESTINAE

Skull Proportions	Index	*Genetta*	*Herpestes ichneumon*	*Xenogale microdon*	*Atilax macrodon*	*Ichneumia leucura*	*Bdeogale nigripes*	Remarks
Total skull length .		82.0	93.0	100.0	100.5	101.0	108.2	
Bizygomatic width. .	I	50 0	51.6	59.0	59.3	55.4	54.5	Slight widening
Max. cran. width. . .	VI	54.7	53.3	63.3	63.9	63.3	63.2	Perceptible increase
Forehead width.	VIII	41.4	48 1	50.0	39.0	55.3	46.4	Variable
Interlacr. width. . . .	II	61.5	66.4	69.7	67.2	64.9	70.4	Moderate increase in *Bdeogale*
Palatal width. . . .	X	73.0	70.7	71.1	75.1	70.5	76.0	Narrow range
Lower dental arch width.	XIV	54.1	58.8	56.4	61.1	54.4	64.0	Increase in *Bdeogale*
Lower jaw spread. . .	XII	49.1	46.0	52.7	48.9	50.7	43.1	Variable
Bicondylar width. . . .								
Olfactory length . . .	III	54.7	55.0	66.7	65.1	68.3	74.0	Marked increase
Orbital length. . . .	IV	34.0	24.2	25.0	26.2	30.8	31 5	Moderate range
Basicranial length. .	XI	47.2	50.3	49.0	42.7	47.0	46.8	Slight
Upper carnassial width	XVIII	70.5	72.2	72.5	90.9	89.7	97.3	
Lower carnassial length.		68.4	67.8	63.2	51.1	53.6	53.3	

The larger-skulled herpestines show marked differences in the dentition in passing from the carnassial *Herpestes* to the "procyonoid" dentition of *Bdeogale*, yet their skull proportions as expressed above are fairly constant except for the variability of forehead width and the increasing length and width of the muzzle. In general their crania are relatively wider than in the Viverrinae and so is the space between the eyes (index II).

Group B

Calogale. A pygmy relative of *Herpestes*, with similar carnassial dentition and narrow skull. The ectotympanic, however, is more inflated ventrally, as is frequently the case in small mammals.

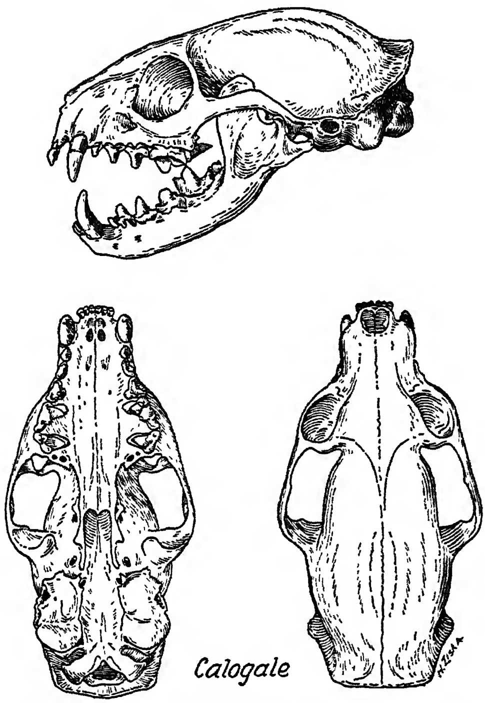


FIG. 28. *Calogale*, three views of skull. × ¾.

Galerella is closely related to *Calogale* but the skull is relatively slightly wider, the muzzle and palate shorter in front. The downward expansion of the ectotympanic is pronounced; the entotympanic, covered by the paroccipital process, is becoming globose posteriorly.

Cynictis penicillata, the bushy-tailed mierkaat, is not properly speaking a mierkaat at all. It shows apparently convergent resemblances to *Galidictis* in the general appearance of the skull, in its general viverrid heritage and in possessing a spout on the external auditory meatus, but differs from *Galidictis* in its many herpestid heritage characters, as follows:

(1) muzzle narrow transversely, much like *Herpestes;*
(2) forehead swollen, much like *Herpestes;*
(3) postorbital bar complete, much like *Herpestes;*
(4) postorbital notch very sharp and directed almost backward;

(5) much swollen ectotympanic, with deep depressed area and conspicuous foramen, just ventral to the external auditory meatus spout;
(6) carotid foramina exceptionally large and piercing the basisphenoid near the midline.

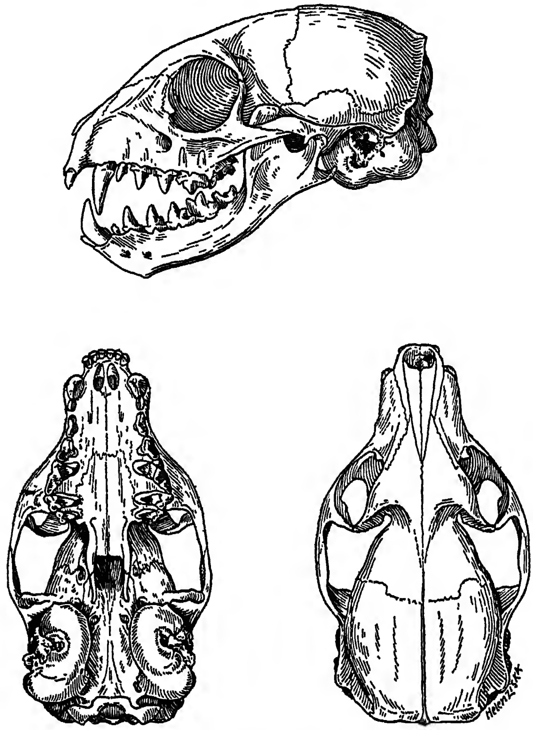


FIG. 29. *Cynictis*, three views of skull. × ¾.

As compared with all other genera of herpestids, the *Cynictis* skull and dentition are nearest to *Galerella*, from which it differs chiefly in the marked widening of the cranium, great inflation of the anterior lobe of the bulla, and increase in size of m^2. The lower jaws and teeth are exceedingly close to *Galerella*, save that m_2 is larger and the lower border of the corpus is less curved downward.

Cynictis differs markedly from *Ichneumia* in its small and far more carnassial cheek teeth. Austin Roberts (cited by Shortridge, under *Cynictis*) suggested that *Cynictis* might be a diurnal derivative of *Ichneumia*, but while the resemblance is close in many details of skull structure, yet *Cynictis* appears to be even closer to *Galerella* but more specialized in its widened cranium and much inflated ectotympanic.

Cynictis approaches *Suricata* in the widening of m^1, m^2, but is much more primitive than the latter in: far less widening of skull, less reduc-

tion of metastyle shear and interdental V on p^4, much less widening of postorbital region of frontals, less shortening of face and of palate. Altogether more primitive, nearer to *Galerella* than to *Suricata.*

From Pocock's figures, the feet of *Cynictis*, although agreeing with *Suricata* in loss of the hallux, differ from those of *Suricata* in: (1) long heel and instep completely hairy (naked in *Suricata*); (2) pollex retained (lost in *Suricata*); (3) manus wider. The feet of *Cynictis* resemble those of *Ichneumia*, especially the manus, but the pes are more advanced in the loss of the hallux. They differ from those of *Galerella* (figured by Allen, 1924, p. 181) in: (1) hallux lost (retained in *Galerella*); (2) heel entirely hairy (only posterior part hairy in *Galerella*); (3) digits all more slender. Hence on all counts *Cynictis* has advanced beyond *Galerella*, often in the general direction of *Suricata*.

Group C

Helogale. A dwarf relative of *Cynictis*, with relatively wider cranium (across squamosals), but narrower forehead; the carnassial p^4 appears relatively smaller than in *Cynictis* and *Galerella*. The feet (Allen, 1924, p. 185) are close to those of *Galerella* but shorter.

Crossarchus. A small relative of *Ichneumia* and possibly also of *Atilax*, pointing toward *Suricata;* forehead, muzzle and cranial contours much as in *Ichneumia;* nose and bony muzzle elongate (see Table 11, index III); eyes small; cheek teeth more advanced toward a pseudo-*Ictops* type; p^4 wide, with short metastyle shear, well developed parastyle and prominent but still slender inner cusp; m^1 wide, short anteroposteriorly; paracone and metacone subequal, buccally projecting parastyle and metastyle; m^2 tending to resemble m^1, larger than in primitive herpestines; m_1 with sharp pricking trigonid, but paraconid and metaconid still well defined, talonid somewhat widened; m_2 small but well developed, trigonid low; paraconid not visible, talonid with basin and hypoconulid rim; p_3 with buccally projecting hypoconid, as in *Atilax;* $p\frac{1}{1}$ absent. Ectotympanic with circular depression medial to spout; orbits directed outward and upward (*cf. Atilax et al.*). Digits 5–5, digits separate, much as in *Atilax* but shorter, heel partly hairy. Tail moderate.

Mungos. A small relative of *Crossarchus*, with incipient shortening of cranium and face. Cheek teeth much as in *Crossarchus* but m^2 smaller. Digits 5–5, somewhat as in *Crossarchus* but pes narrower, with hallux farther from the distal end.

Suricata. The slender-tailed mierkaat. Cranium much shortened and widened, especially posteriorly. Face short, sharply deflected on cranium. Cheek teeth pseudo-*Ictops*-like: p^4 wide with much reduced

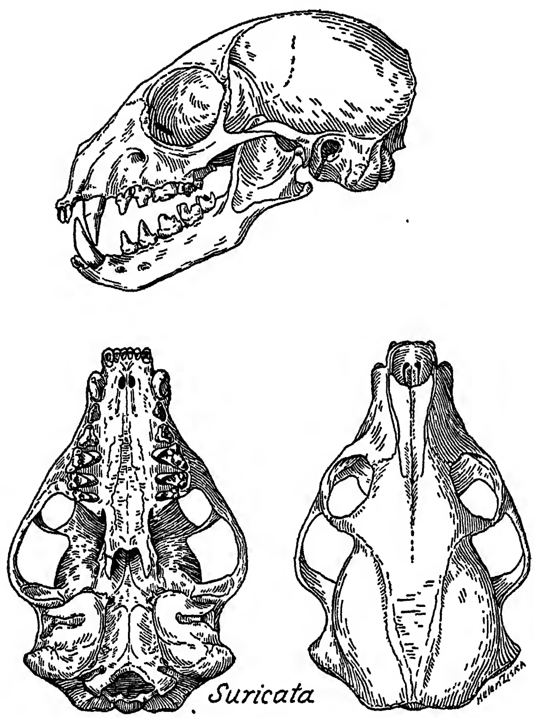


FIG. 30. *Suricata*, three views of skull. × ¾.

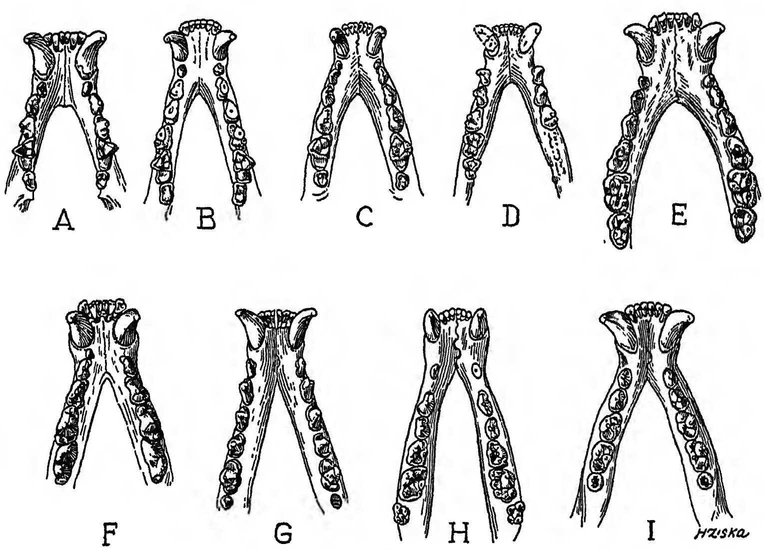


FIG. 31. Series of lower jaws. Scales various.

A. Galidictis. B. Herpestes ichneumon. C. Calogale. D. Suricata. E. Bdeogale sp. *F. Cryptoprocta. G. Nandinia. H. Arctogalidia. I. Arctictis.*

metastyle; m^1 wide, acute inner carnassial angle; subequal paracone and metacone, a metastyle cusp; m_1 with small connate paraconid and metaconid; p^1_1 absent; upper canine long, slender, nearly straight. Ectotympanic with conspicuous transverse slot. Digits 4–4. Tail slender.

TABLE 11

THE SMALLER HERPESTINES: *Calogale* TO *Cynictis*, *Helogale* TO *Suricata*

Skull Proportions	Index	Group B			Group C				Remarks
		Genetta	*Calogale*	*Cynictis*	*Helogale*	*Mungos*	*Crossarchus*	*Suricata*	
Total skull length...		82.0	60.9	64.0	50.3	61.0	74.5	57.2	All pygmies
Bizygomatic width..	I	50.0	52.9	60.6	56.7	60.7	61.7	75.5	Widening in B and C
Max. cran. width....	VI	54.7	58.3	65.1	73.5	68.3	70.5	77.5	Marked widening in C
Forehead width.....	VIII	41.4	48.1	52.1	45.6	48.9	46.1	71.0	Marked widening in C
Interlacr. width.....	II	61.5	72.2	75.9	78.7	86.0	71.2	72.6	Wide, especially in *Mungos*
Palatal width.......	X	73.0	77.2	78.0	79.7	85.2	81.2	82.9	Marked widening in C
Lower dent. arch width............	XIV	54.1	55.4	54.8	64.2	70.4	57.6	63.9	Some widening in C
Lower jaw spread...	XII	49.1	49.8	52.7	61.2	59.1	49.5	63.2	Extreme in *Suricata*
Bicondylar width....									
Olfactory length	III	54.7	48.5	48.8	47.9	48.8	69.3	43.0	*Crossarchus* has very large nose
Orbital length.......	IV	34.0	24.9	31.9	41.2	22.4	26.1	33.0	In *Helogale*, short brain makes IV high
Basicranial length...	XI	47.2	47.8	44.6	44.1	43.9	48.4	33.8	Marked shortening in *Suricata*
Upper carnassial width............		70.5	77.4	86.9	97.2	102.0	82.1	120.0	
Lower carnassial length...........		68.4	61.8	55.2	51.2	51.1	60.7	51.1	

The pygmy herpestines show a moderate to pronounced increase of bizygomatic index, a marked widening of the cranium, of the forehead (culminating in *Cynictis* of Group B and *Suricata* of Group C). Muzzle and palate widening (especially in Group C) and in the latter also the lower dental arch. The olfactory length displays a sudden increase in *Crossarchus*.

SUMMARY AND CONCLUSIONS

Our studies on the skulls and dentitions of the civets and their allies have led to the suggestions or conclusions listed below.

1. The subordinal terms Creodonta, Fissipedia, and even Pinnipedia, although here (page 313) retained in square brackets on account of their historical interest, are more or less arbitrary or artificial as-

semblages. A more natural division of the order would be into the infra-orders Procreodi, Acreodi, Pseudocreodi, and Eucreodi of Matthew, based on the different positions in the cheek teeth series of the most highly developed sectorial tooth.

2. The name Eucreodi, although restricted by Matthew to the "Creodonta Miacidae," might well be expanded to include also both the Fissipedia and the Pinnipedia, since there is evidence (direct or indirect) that *all* the families of modern Carnivora were derived from one common ancestral stock in which the principal sectorial teeth were the fourth upper premolars and the first lower molars $\left(\frac{p^4}{m_1}\right)$.

3. The Eocene Palaeonictidae, which have the *principal* carnassial blades on p^4 and m_1, but also have somewhat smaller carnassials on m^1 and m_2, are transitional in this and other respects between the "Pseudocreodi" and the "Eucreodi" and for this reason are here assigned to a new infraorder, the Amphicreodi.

4. Since the principal carnassials are located in the Pseudocreodi behind p^4, m_1, in the Amphicreodi on both p^4, m_1 and m^1, m_2, and in the Eucreodi wholly on p^4, m_1, it seems possible that the varying location of the chief carnassials is due to a shift in maximum growth gradients from $\frac{m^1}{m_2}$ to $\frac{p^4}{m_1}$.

5. Matthew's American Eocene "Creodonta Miacidae" have the entire dentition closely resembling that of the primitive Viverridae, and among the European lower Oligocene Miacidae there are forms which can barely if at all be distinguished from the Viverridae in their dentition, as noted by Teilhard de Chardin.

It was supposed by Matthew that the absence of an ossified bulla was characteristic of the American Eocene Miacidae, but whether or not this was due merely to imperfect preservation of the fossils, it is a fact that some of the lower Oligocene European "Miacidae" without bony bullae are essentially similar in many other features to the Viverridae, which for the most part have fully ossified bicameral bullae. For these and other reasons we regard Matthew's "Miacidae" as being nothing but primitive forerunners and ancestors of the later civets and their allies. Accordingly we propose to merge the family Miacidae with the Viverridae, retaining the latter name for the whole assemblage.

6. The subfamilies and families of the aeluroid Carnivora recognized by Pocock have every appearance of being compact "natural" groups but we conclude that by his rigid use of arbitrarily chosen "leading characters" in defining his subfamilies and families he has separated:

Prionodon too widely from the true Viverrinae;
Proteles too widely from the Hyaenidae;
Cryptoprocta too widely from the Felidae;
Nandinia too widely from the Paradoxurinae;
Arctogalidia too widely from the Paradoxurinae;
Fossa too widely from the Hemigalinae;
Cynogale too widely from the Hemigalinae;
Eupleres too widely from the Hemigalinae;
Galidictis too widely from the Fossinae.

7. In order not to lose the advantages gained by Pocock's sharp analysis and at the same time to indicate the more remote phylogenetic relationships of each subfamily, which relationships are largely obscured by his system, we have grouped his subfamilies into larger assemblages called sections; these we regard as the primary divisions of the family; and we have used the termination *ida* to designate them.

8. We therefore divide the family Viverridae into the sections Viverrida, Paradoxurida, Hemigalida, Galidictida; the family Hyaenidae into the sections Protelida, Hyaenida; the family Felidae into the sections Cryptoproctida, Felida and Machaerodontida, all containing the subfamilies listed on page 320 above.

9. In contrast to some older theories, it now seems highly probable that at the structural horizon represented by the existing aeluroid Carnivora, the most primitive types of cheek teeth are not those of the omnivorous or procyonine type realized in such forms as *Bdeogale*, but those which may be called the central carnassial type illustrated by the Paleocene *Propappus* and *Didymictis* and to a somewhat less extent by the modern *Viverra* and *Herpestes*. This central carnassial (Fig. 4) type is characterized as follows:

(*a*) p^4 possesses a large and unreduced, notched paracone-metastyle blade, which is directed obliquely backward and somewhat outward;

(*b*) against the inner side of this blade shears the outer side of the paraconid-protoconid blade of m_1, consisting essentially of two triangular blades separated by a deep notch (Fig. 5);

(*c*) the inner basal cusp of p^4 is of small size and is displaced to the antero-internal corner of the tooth, so as to allow room for the paraconid blade of m_1 and to act as a stop or holdfast that presses the food against that blade;

(*d*) the first upper molar is widened transversely and its anterior margin serves as an antagonist for the notched protoconid-metaconid blade of m_1;

(*e*) the talonid or heel of m_1, while smaller than the trigonid, is still functional: it receives the stout internal cusp or protocone of m^1 and its posteroexternal cusp or hypoconid shears between the inner faces of the para- and metacone of m^1;

(*f*) the "outer carnassial angle" (defined as in Fig. 32) between the outer margins of p^4 and m^1, which is about 130° in the central carnassial type (*Viverra*), approaches a right angle in the extreme carnassial stage;

(*g*) the "inner carnassial angle" (defined as in Fig. 32), which is about 37° in the central carnassial type, opens out to 60° or more in extreme carnassial types;

(*h*) the protoconid-paraconid notched blade of m_1 in the central carnassial type, which makes an angle of about 45° degrees with a line drawn from the anterior basal point of p_3 to the postero-median point of m_2, becomes essentially parallel to a similar line (drawn, however, to the postero-median point of m_1) in the advanced carnassial stage (Fig. 5B).

10. From this central carnassial type of p^4 and m_1 one line of specialization leads, as above noted, to the extreme carnassial stage, with great emphasis of the paracone-metastyle blade of p^4, reduction and anterior displacement of the internal basal cusp of that tooth, together with enlargement of the protoconid-paraconid notched shear of m_1, with reduction and eventual disappearance of its metaconid, reduction and loss of its talonid, and anteroposterior alignment of its paraconid-protoconid blade.

11. As noted above, the central carnassial type is retained with slight modifications in the subfamilies Viverrinae, Prionodontinae, Galidictinae, Herpestinae.

12. More advanced carnassial specialization occurs in *Cryptoprocta* and the Hyaenidae, while the most advanced stage is reached by the Felidae.

13. The feeble peg-like cheek teeth of *Proteles* appear to have arisen by rapid degeneration from the primitive hyaenid stage (p. 332).

14. The opposite line of specialization, culminating in *Bdeogale* of the Herpestidae) gradually transforms the shear-toothed central carnassial type into the blunt-toothed crowns of the procyonine type, in which the metastyle blade of p^4 is reduced, the internal basal cusp of that tooth secondarily enlarged; a pseudo-hypocone appears on the postero-internal margin of the crown; the first and second upper molars become secondarily enlarged, with blunt cusps, the "outer carnassial angle" disappears, opening out to about 160°; meanwhile the protoconid-paraconid blade of m_1 has disappeared, as all the cusps of this tooth have become

roundly conical; an accessory cusp appears at the antero-external base of the protoconid; the talonid of this tooth has become widened to receive the enlarged protocone of m^1; m_2 has enjoyed a secondary increase in size, especially in length.

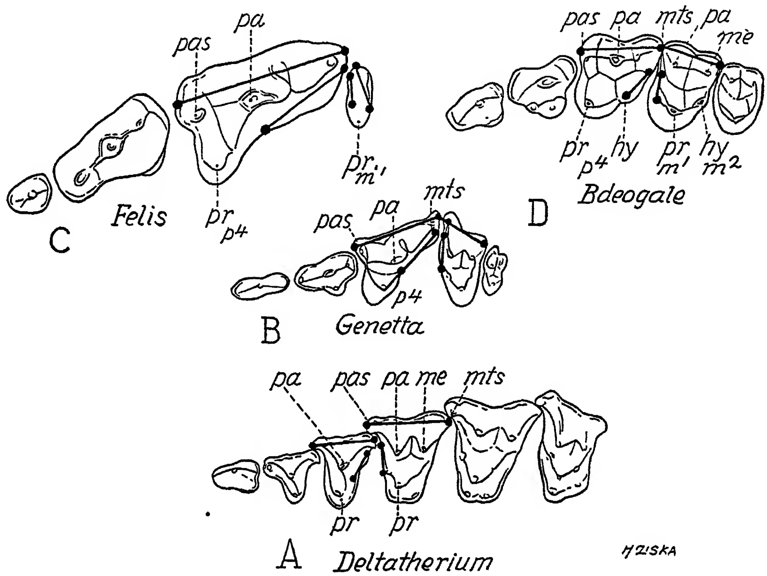


FIG. 32. Relations of the outer and inner carnassial angles in primitive carnassial and procyonoid dentitions.

A. Deltatherium. B. Genetta. C. Felis. D. Bdeogale.

15. Starting from the central carnassial type (largely preserved in *Nandinia*), the palm civets went through the initial steps toward the procyonine stage; then the binturongs and *Arctogalidia* suffered a marked reduction in the size of the cheek teeth, which are left with feeble conical cusps.

16. In the fanaloka (*Fossa*) a modification of the primitive carnassial type is expressed in the outward deflection of the metastyle shear of p^4, in the widening of the talonid basin of m_1, m_2, and in the reduction and twinning of the para- and metaconids of m_2; here also p_2, p_3, p_4, tend to be compressed, with incipiently serrated crowns.

17. In *Hemigale* and *Diplogale* the upper carnassial tends to become like the first upper molar, with roundly conical cusps; m^1, m^2, acquire conspicuous central basins; likewise m_1 loses its carnassial character and m_2 becomes secondarily enlarged.

18. *Cynogale* is a blunt-toothed derivative of *Hemigale*, with still more compressed, piscivorous premolars.

19. *Chrotogale* is a weak-toothed derivative of *Diplogale,* with small cheek teeth spaced in front, peculiar, kangaroo-like upper incisors and partly procumbent lower incisors.

20. *Eupleres* is a degenerate derivative of *Chrotogale,* with much enfeebled cheek teeth, whose secondarily sharp points give a wholly false impression of primitiveness. Its front teeth are derivable by degeneration from those of *Chrotogale* and the same is true of its deciduous teeth. A comparison of the skull characters of these genera gives strong support to this conclusion (pp. 362, 363).

21. The Galidictines are pgymy derivatives of the primitive carnassial type, with relatively wide p^4, m^1, and sharp outer carnassial angle. They may be pygmy relatives of *Fossa.* *Galidictis* has relatively huge lower canines and a somewhat skunk-like habitus of the skull. Resemblances to the Herpestidae may be due in part to parallelism.

22. Within the Herpestidae the first series of genera (*Herpestes* to *Bdeogale*) shown in Pl. VI affords a complete demonstration of the transformation of the typical carnassial into a secondarily procyonine stage (in the manner already noted); the second and third series of genera (*Calogale* to *Cynictis, Helogale* to *Suricata*) illustrate the transformation of the central carnassial type into a secondarily insectivorous or ictopsoid type, involving the marked anteroposterior shortening of the crowns of p^4, m^1, m^2, the approximation in form of p^4 and of m^1 toward the narrow V-form seen in *Ictops,* with the loss of carnassial form on m_1 and the quite secondary evolution of a high pricking trigonid with twinned para- and metaconids on m_1, m_2.

23. The terminal members of the second and third series (*Cynictis* and *Suricata*) resemble each other in many habitus characters of the skull and dentition, and even the living animals run about, stand up on their hind legs and lean back upon their tails in much the same way. Our studies of the skull and dentition indicate clearly that *Cynictis* and *Suricata* have paralleled each other in the features noted above and that they are shown to belong to different series by the possession of contrasting "heritage characters," especially in the bullae. Nevertheless we regard this case as a good illustration of the principle that parallelism or independent acquisition of similar characters in related phyla after the separation of the phyla is in itself evidence of relationship.

24. Our studies of the compound auditory bullae of the families and subfamilies of the viverrids and their allies tend to confirm the value of this region as usually yielding good diagnostic characters, at least for subfamilies. Although Pocock has shown that there are noticeable individual variations in detail and in the proportions of the two com-

ponent parts of this region within certain genera, we have found that among genera of the same family the auditory bullae are often more conservative than the dentition in retaining old heritage characters.

25. It is well known that the tympanic (ectotympanic) bone in such primitive mammals as the opossum (*Didelphis*) is a horseshoe-shaped ring supporting the drum membrane and this condition is found in the foetal stages of modern Carnivora. The entotympanic has apparently been developed independently in the Insectivora and Carnivora. It appears to represent an epiphysis formed around the periphery of the ectotympanic. Its expansion seems to have occurred *pari passu* with that of the posterior swelling of the tubo-tympanal canal.

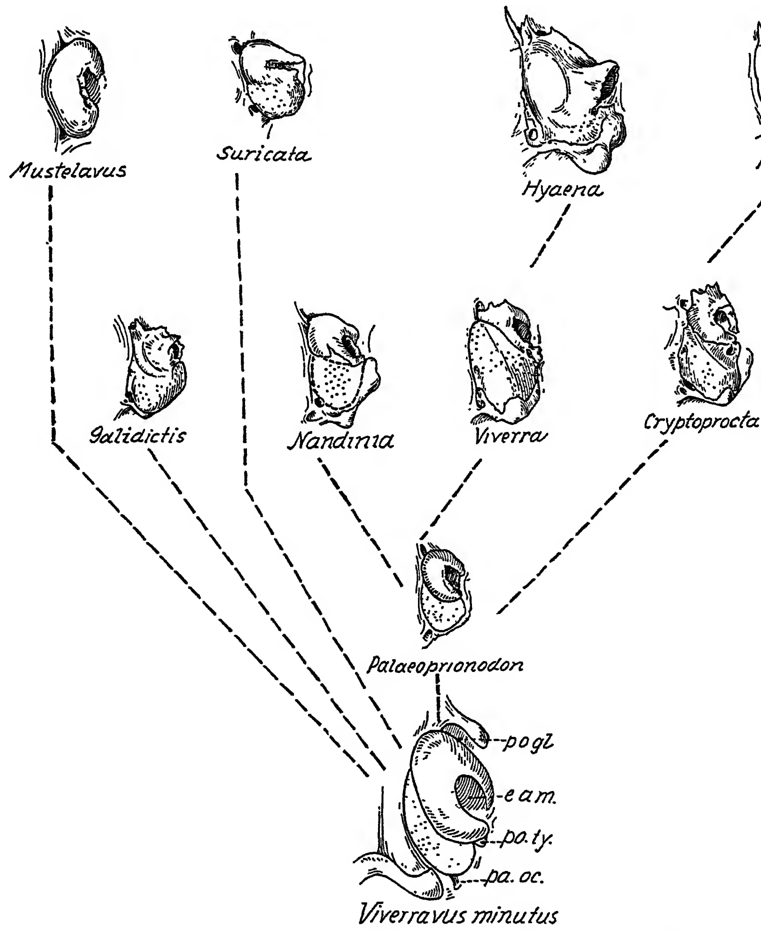


FIG. 33. Evolution of the compound bulla in Aeluroid Carnivora. The entotympanic is stippled, the ectotympanic plain.

The evolution and adaptive radiation of the auditory bullae in the viverrids and related groups may be summarized as follows:

(*a*) The primitive condition (for aeluroid Carnivora) is apparently represented in *Viverravus*, as figured by Teilhard de Chardin (1915, Pl. I). Here the tympanic (ectotympanic) bone is large and well ossified; the entotympanic, if present, may have been only a slight epiphysial margin on the medial side of the ectotympanic.

(*b*) In *Palaeoprionodon* (as figured by Teilhard de Chardin, Pl. IX) the ectotympanic is restricted to the region around the bony external auditory meatus; space for the entotympanic, which is believed by Teilhard and others to have been membranous, is postero-internal to the ectotympanic; already this chamber had expanded postero-externally to such a degree as to cause the expansion of the paroccipital process of the exoccipital.

(*c*) In the Viverrinae (see p. 326) the entotympanic chamber is elongated anteroposteriorly and widened postero-externally; the antero-internal apex of the entotympanic never reaches forward as far as the anterior edge of the ectotympanic.

(*d*) In *Cryptoprocta* the entotympanic is only moderately enlarged and its antero-internal angle extends only to about the mid-point of the medial border of the ectotympanic.

(*e*) In the cats the entotympanic is well inflated and its antero-internal angle extends far forward, to a point antero-internal to the ectotympanic portion of the compound bulla.

(*f*) In the hyaenas the ectotympanic has expanded, the entotympanic has retreated and forms only a small portion of the posterior part of the bulla.

(*g*) In *Proteles* the bulla is inflated vertically; most of it seems to belong to the ectotympanic but there is a small chamber in the rear which probably represents the entotympanic. Thus *Proteles* in the construction of its bulla is intermediate between the Viverrinae and the Hyaenidae.

(*h*) The entotympanic bulla in *Fossa* is rather small and subglobose posteriorly; it is embraced by the concave shell of the paroccipital region; the moderate ectotympanic is not much inflated.

(*i*) In *Nandinia* the entotympanic chamber has become membranous but traces of its bony character may be seen in some specimens. From other evidence we conclude that the foregoing condition is not primitive, as has been generally assumed, but quite secondary.

(*j*) In the "small-toothed palm civets" (Arctogalinae) the entotympanic bulla lacks the prominent vertical ridge which is conspicuous

in the Paradoxurinae, but this and other differences hardly obscure the relatively close relationship with the Paradoxurinae.

(*k*) The auditory bulla of *Paradoxurus* is "conical, broad, and truncated behind, pointed in front and rather compressed on the sides, which meet in a ridge" (Flower). Such conditions may have been derived from those seen in *Fossa*.

(*l*) In the Hemigalida (Fossinae, Hemigalinae, Cynogalinae, Euplerinae) the ectotympanic is but little or not inflated, enclosing the auditory meatus in a prominent oval, directed obliquely upward and backward; entotympanic chamber produced anterointernally (except in *Fossa*), well inflated and overlapping ectotympanic anterointernally. Derivation from the stem of the Paradoxurinae not far from *Fossa* is indicated.

(*m*) In the Galidictinae the ectotympanic is somewhat inflated and bears a prominent spout; the entotympanic is inflated transversely, flat ventrally. Derivation from near *Fossa* seems likely.

(*n*) In the Herpestidae the ectotympanic has a more or less spout-like external auditory meatus, the entotympanic is typically inflated and globose. The family may be divided into three groups: A (*Herpestes* to *Bdeogale*), ectotympanic less inflated ventrally than entotympanic; B (*Calogale* to *Cynictis*), ectotympanic inflation extending ventrally below level of entotympanic inflation; C (*Helogale* to *Suricata*), ectotympanic becoming depressed and widened, finally with wide transverse groove or slot.

(*o*) The dominance of the ectotympanic in the Mustelidae and their tendency to form an ectotympanic spout may be due either to direct inheritance from a form like *Plesictis*, or it may be derived from a condition such as that in *Cynictis* of the Herpestidae, in which the ectotympanic has become secondarily predominant over the entotympanic.

26. The feet of the primitive Viverridae were probably pentadactylate, of subcursorial, digitigrade type, with retractile claws and narrow tarsus. Spreading plantigrade hands and feet have, we infer, been acquired secondarily by arboreal types. The Herpestidae often use their small feet for scratching and digging; their claws are long, not much curved, sometimes tending to lose the pollex and hallux.

27. In general our studies incline us to question current applications of the doctrine of irreversibility of evolution to the major classification and phylogeny of the Carnivora, and to suggest that among the aeluroid Carnivora, as well as in many other groups, one can find many gradations in any particular structure or group of structures. Those who have repudiated what is often called "end-on evolution" have over-

looked, we believe, the fact that structural ancestors and their derivatives of varying degrees of remoteness may often be found living in the same geological epoch.

Literature Consulted

Allen, J. A. 1924. Carnivora Collected by the American Museum Congo Expedition. *Bull. Amer. Mus. Nat. Hist.*, XLVII, Art. 3: 73–281, Pls. VI–LXXVIII, 67 Figs.

Andrews, C. W. 1906. A Descriptive Catalogue of the Tertiary Vertebrata of the Fayûm, Egypt. London: Printed by Order of the Trustees of the British Museum. 324 pp., 26 Pls., 97 Figs.

Beddard, Frank Evers. 1902. Mammalia. Cambridge Natural History (Ed. S. F. Harmer and A. E. Shipley), X, 605 plus xii pp., 285 Figs.

Blainville, H. M. D. de. 1839–64. Ostéographie des Mammifères. Atlas. 117 Pls. Paris.

Blanford, W. T. 1878. On Some Mammals from Tenasserim. *Journ. Asiatic Soc. Bengal*, XLVII, Part 2, No. 10: 150–167, Pls. VI, VII, VIII.

Boule, M., and Jean Piveteau. 1935. Les Fossiles. Éléments de Paléontologie. Masson et Cie. Paris. 899 pp., 1330 Figs.

Carlsson, Albertina. 1902. Ueber die systematische Stellung von *Eupleres goudoti*. *Zool. Jahrb.*, XVI: 217–242, Pls. X, XI.

——. 1910. Die genetischen Beziehungen der madagassischen Raubtiergattung *Galidia*. *Zool. Jahrb.*, XXVIII: 559–602, Figs. A–V.

——. 1911. Über *Cryptoprocta ferox*. *Zool. Jahrb.*, XXX: 419–470, Pls. XII–XIV.

——. 1920. Über *Arctictis binturong*. *Acta Zoologica*, 1920: 337–380, Pls. I–III.

Clark, John. 1937. The Stratigraphy and Paleontology of the Chadron Formation in the Big Badlands of South Dakota. *Ann. Carnegie Mus.*, XXV: 261–350, Pls. XXI–XXVI, 12 Figs.

Cope, E. D. 1884. The Vertebrata of the Tertiary Formations of the West. U. S. Geological Survey of the Territories, 1883, III, Book I (2 vols.). 1009 + xxxiv pp., Pls. I–LXXVa. U. S. Geol. Surv. Washington. 1884.

Denison, Robert Howland. 1938. The Broad-skulled Pseudocreodi. *Ann. N. Y. Acad. Sci.*, XXXVII, Art. 3: 163–256, 32 Figs.

Filhol, M. H. 1879. Étude des Mammifères fossiles de Saint-Gérard le Puy (Allier). Part I. Biblioth. d. l'École des Hautes Études, Section d. *Sci. Nat.*, XIX: 1–252, 30 Pls.

——. 1884. Mémoires sur quelques Mammifères fossiles des Phosphorites du Quercy. Toulouse, 1884 (1882). 46 pp., 13 Pls.

Flower, W. H. 1885. An Introduction to the Osteology of the Mammalia. (3rd. Ed.) London. Macmillan and Co. 382 pp., 132 Figs.

Frick, Childs. 1926. The Hemicyoninae and an American Tertiary Bear. *Bull. Amer. Mus. Nat. Hist.*, LVI: 1–119, 63 Figs.

——. 1937. Horned Ruminants of North America. *Bull. Amer. Mus. Nat. Hist.*, LXIX: 1–669 pp., 103 Figs.

Gervais, Paul. 1874. Dentition et squelette de l'Euplère de Goudot. *Journ. d. Zool.*, III: 237–251, Pl. VII.

Gray, J. E. 1870. Description of an Adult Skull of *Eupleres goudoti*. *Proc. Zool. Soc. London*, Dec. 6, 1870: 824–828, Pl. LI, 2 Figs.

Gregory, William K. 1933. Nature's Wild Dog Show. *Bull. N. Y. Zool. Soc.*, XXXVI: 83–96, 23 Pls., 4 Figs.

——. 1934. A Half Century of Trituberculy: the Cope-Osborn Theory of Dental Evolution. *Proc. Amer. Philos. Soc.*, LXXIII: 169–317, 71 Figs.

——. 1935. Polyisomerism and Anisomerism in Cranial and Dental Evolution among Vertebrates. *Proc. Nat. Acad. Sci.*, XX: 1–9.

——. 1936. On the Phylogenetic Relationships of the Giant Panda (*Ailuropoda*) to Other Arctoid Carnivora. *Amer. Mus. Novitates*, No. 878, 29 pp., 22 Figs.

Hodgson, B. H. 1842. On a New Species of *Prionodon*, *P. Pardicotor nobis*. *Calcutta Journ. Nat. Hist.*, II: 57–60, Pl. I.

Hollister, N. 1918. East African Mammals in the United States National Museum. Part I. Insectivora, Chiroptera, and Carnivora. Smiths. Inst., U. S. Nat. Mus., Bull. 99: 1–194, 55 Pls.

JEPSEN, G. L. 1930. Stratigraphy and Paleontology of the Paleocene of Northeastern Park County, Wyoming. *Proc. Amer. Philos. Soc.*, LXIX, No. 7: 463–528.

LECHE, W. 1909. Zur Frage nach der stammesgeschichtlichen Bedeutung des Milchgebisses bei den Säugetieren. I. Mitteilung. *Zool. Jahrb.*, XXVIII: 449–456, Pl. 9.

——. 1915. Zur Frage nach der stammesgeschichtlichen Bedeutung des Milchgebisses bei den Säugetieren. II. Viverridae, Hyaenidae, Felidae, Mustelidae, Creodonta. *Zool. Jahrb.*, XXXVIII: 275–370, 126 Figs.

LOOMIS, F. B. 1932. The Small Carnivors of the Miocene. *Amer. Journ. Sci.*, XXIV: 316–329, 10 Figs.

LYDEKKER, RICHARD. The Carnivores: Order Carnivora. In New Natural History [Ed. R. Lydekker], I, Sect. II, Chap. XIII: 349–576; II, Sect. III, Chap. XVI, XVII, XVIII: 1–101. Illustrated. New York. Merrill and Baker.

MATTHEW, W. D. 1899. The Ancestry of Certain Members of the Canidae, the Viverridae and Procyonidae. (With J. L. Wortman.) *Bull. Amer. Mus. Nat. Hist.*, XII: 109–138, Pl. VI, 10 Figs.

——. 1901. Additional Observations on the Creodonta. *Bull. Amer. Mus. Nat. Hist.*, XIV: 1–38, 17 Figs.

——. 1909. The Carnivora and Insectivora of the Bridger Basin, Middle Eocene. *Mem. Amer. Mus. Nat. Hist.*, IX: 291–567, Pls. XLIII–LII, 118 Figs.

——. 1915. A Revision of the Lower Eocene Wasatch and Wind River Faunas (With Walter Granger). Part I. Order Ferae (Carnivora), Suborder Creodonta (By W. D. Matthew). *Bull. Amer. Mus. Nat. Hist.*, XXXIV: 1–103 pp., 87 Figs.

——. 1928. The Evolution of the Mammals in the Eocene. *Proc. Zool. Soc. London*, Part 4 (1927): 947–985, 16 Figs.

——. 1937. Paleocene Faunas of the San Juan Basin, New Mexico [Ed. by Walter Granger, William K. Gregory, Edwin H. Colbert]. *Trans. Amer. Philos. Soc.*, N. S., XXX: 1–510, 65 Pls., 85 Figs.

MILNE EDWARDS, A., AND ALFR. GRANDIDIER. 1867. Observations anatomique sur quelques mammifères de Madagascar. I. De l'organisation du *Cryptoprocta ferox*. *Ann. d. Sci. nat.*, (5), VII: 314–338, Pls. 7–10.

MIVART, ST. GEORGE. 1882*a*. On the Classification and Distribution of the Aeluroidea. *Proc. Zool. Soc. London*, 1882, No. 1: 135–208, 15 Figs.

——. 1882*b*. Notes on Some Points in the Anatomy of the Aeluroidea. *Proc. Zool. Soc. London*, 1882: 459–520, 12 Figs.

OGILBY, W. 1835. On the Characters and Description of a New Genus of Carnivora Called *Cynictis*. *Trans. Zool. Soc. London*, I: 29–34, Pl. III.

OSGOOD, WILFRED H. 1932. Mammals of the Kelley-Roosevelts and Delacour Asiatic Expeditions. *Field Mus. Nat. Hist., Publ.* 312, *Zool. Ser.*, XVIII: 193–339, Pls. X, XI, 2 Figs.

POCOCK, R. I. 1915*a*. On Some of the External Characters of *Cynogale bennettii* Gray. *Ann. Mag. Nat. Hist.*, (8), XV: 351–360, Pl. XIV.

——. 1915*b*. On the Species of the Mascarene Viverrid *Galidictis*, with the Description of a New Genus and a Note on "*Galidia elegans*." *Ann. Mag. Nat. Hist.*, (8), XVI: 113–124, Pl. VII.

——. 1915*c*. On Some of the External Characters of the Palm-Civet (*Hemigalus derbyanus* Gray) and its Allies. *Ann. Mag. Nat. Hist.*, (8), XVI: 153–162, Pl. VIII.

——. 1915*d*. On the Feet and Glands and Other External Characters of the Paradoxurine Genera *Paradoxurus*, *Arctictis*, *Arctogalidia* and *Nandinia*. *Proc. Zool. Soc. London*, September, 1915: 387–412, 10 Figs.

——. 1915*e*. On Some of the External Characters of the Genus *Linsang*, with Notes upon the Genera *Poiana* and *Eupleres*. *Ann. Mag. Nat. Hist.*, (8), XVI: 341–351, Pls. XII, XIII.

——. 1915*f*. On Some External Characters of *Galidia*, *Galidictis*, and Related Genera. *Ann. Mag. Nat. Hist.*, (8), XVI: 351–356, Pls. XIV, XV.

——. 1916*a*. The Tympanic Bulla in Hyaenas. *Proc. Zool. Soc. London*, February 22, 1916: 303–307, 2 Figs.

——. 1916*b*. On the External Characters of the Mongooses (Mungotidae). *Proc. Zool. Soc. London*, June, 1916: 349–374, 10 Figs.

——. 1916*c*. On Some of the External Characters of *Cryptoprocta*. *Ann. Mag. Nat. Hist.*, (8), XVII: 413–425, 4 Figs.

——. 1921*a*. The Auditory Bulla and Other Cranial Characters in the Mustelidae. *Proc. Zool. Soc. London*, XXXI: 473–486, 5 Figs.

——. 1921*b*. On the External Characters and Classification of the Mustelidae. *Proc. Zool. Soc. London*, LIII: 803–837, 13 Figs.

——. 1929*a*. The Structure of the Auditory Bulla in the Procyonidae and the Ursidae, with a Note on the Bulla of *Hyaena*. *Proc. Zool. Soc. London*, Part 4, (1928) 1929: 963–974, 8 Figs.

——. 1929*b*. Carnivora. Encyclopaedia Britannica, 14th ed., IV: 896–900.

——. 1933*a*. The Civet-Cats of Asia. *Journ. Bombay Nat. Hist. Soc.*, XXXVI: 423–449, 4 Figs.

——. 1933*b*. The Rarer Genera of Oriental Viverridae. *Proc. Zool. Soc. London*, 1933: 969–1035, 3 Pls., 7 Figs.

——. 1937. The Mongooses of British India, Including Ceylon and Burma. *Journ. Bombay Nat. Hist. Soc.*, XXXIX, No. 2: 211–245, 2 Figs.

Rand, A. L. 1935. On the Habits of Some Madagascar Mammals. *Journ. Mammalogy*, XVI, No. 2: 89–104.

Schlosser, Max. 1899. *Parailurus anglicus* und *Ursus Böckhi* aus den Ligniten von Baróth-Köpecz. Mittheil. a. d. Jahrb. d. Königl. Ungarisch. Geol. Anstalt., Bd. XIII, Heft 2: 1–31, Pls. X, XI, XII.

Schwarz, Ernst. 1935. On a New Mongoose from the Cameroons. *Ann. Mag. Nat. Hist.*, XV, (10), No. 25: 300–301.

Scott, W. B., and Glenn L. Jepsen. 1936. The Mammalian Fauna of the White River Oligocene. Part I. Insectivora and Carnivora. *Trans. Amer. Philos. Soc.*, N. S., XXVIII: 1–153, 22 Pls., 7 Figs.

Shortridge, G. C. 1934. The Mammals of Southwest Africa. Vol. I. London. William Heineman. 437 pp. [*Genetta albiventris*, opp. p. 115.]

Simpson, G. G. 1931. A New Classification of Mammals. *Bull. Amer. Mus. Nat. Hist.*, LIX, Art. V: 259–293.

——. 1937. The Fort Union of the Crazy Mountain Field, Montana, and its Mammalian Faunas. *U. S. Nat. Mus., Bull.* 169: 1–287, 10 Pls., 80 Figs.

Teilhard de Chardin, P. 1915. Les Carnassiers des Phosphorites du Quercy. *Ann. d. Paléont.*, IX, (1914–1915), Fasc. III–IV: 103–193, Pls. XII (I)–XX (IX), 13 Figs.

Thomas, Oldfield. 1892. On Some Mammals from Mount Dulit, North Borneo. *Proc. Zool. Soc. London*, March 15, 1892, No. 2: 221–227, Pl. XIX.

——. 1912. Two New Genera and a New Species of Viverrine Carnivora. *Proc. Zool. Soc. London*, March 19, 1912: 498–503, Figs. 61, 62, 63.

——. 1927. The Delacour Exploration of French Indo-China Mammals. *Proc. Zool. Soc. London*, 1927, No. 3: 41–58, 2 Pls.

Wortman, J. L. 1901–1903. Studies of Eocene Mammalia in the Marsh Collection, Peabody Museum. *Am. Journ. Sci.*, 1901–1903, XI, XII, XIII, XIV.

THE DISTRIBUTION OF PHOSPHATES IN THE CHESAPEAKE BAY *

CURTIS L. NEWCOMBE † AND ANDREW G. LANG

University of Maryland and Chesapeake Biological Laboratory

(Communicated by Edwin G. Conklin)

Abstract

The vertical, horizontal, seasonal, and hourly distribution of inorganic phosphorus in Chesapeake Bay during the period June, 1938–April, 1939, is described. Special attention has been given to the concentrations in the Patuxent River as compared with those in the Bay waters. It is shown that the phosphorus content of the surface water of the Bay during summer was rather uniform whereas in the subsurface layers there was a progressive increase from the mouth of the Bay to the headwaters. In this season, notably higher concentrations characterized the subsurface levels. Highest concentrations were present in summer and the lowest values obtained during October to March, inclusive. Hourly variations were most conspicuous near the hours of sunset and sunrise, being characterized by a pronounced increase shortly after dark and a marked decrease following sunrise. These and other variations are interpreted in light of the metabolism of the diatom population and the turbidity conditions that are prominently developed in the Chesapeake area.

Introduction

The importance of inorganic phosphates in sustaining the phytoplankton of oceans, lakes, and rivers has been pointed out by many investigators. In surface waters of the sea the concentration of dissolved inorganic phosphates is characteristically greatest during the winter months, becoming lessened through the spring, and attaining minimum values in the summer. No information has been available concerning phosphates in the Chesapeake Bay and, consequently, the rôle these substances may play in the economy of the Bay life has been undetermined. A study of inorganic phosphates was therefore included in a program of research bearing on the oxygen-depleted waters of this Bay (Newcombe and Horne, 1938) in an effort to develop a fuller understanding of phosphate characteristics that obtain, namely, the magnitude of phosphate concentrations that prevail; the hourly, daily, and seasonal variations that occur; the vertical distribution; and to what extent the factors operating in the ocean may be significant in the waters of the Bay. Because of the salinity characteristics encountered it was furthermore desirable to develop a method of phosphate estimation suitable to the area.

* Contribution No. 28 of the Chesapeake Biological Laboratory.

† For assistance in the pursuit of the problem, this author is indebted to the American Philosophical Society for a Research Grant.

The problem has been supported through the cooperation of the Chesapeake Biological Laboratory, the American Philosophical Society and the University of Maryland. During its progress, especially valuable assistance has been given by Dr. R. V. Truitt, Director of the Laboratory, to whom the writers are deeply indebted. The efforts of F. M. McNall in contributing to improvements in the method of analysis and for other assistance in the course of the work during the summer is similarly acknowledged.

Method of Analysis

Many variations of Deniges' ceruleomolybdate method of estimating inorganic phosphates are in use in different laboratories. In the present work the molybdate reagent consisted of 25 grams of ammonium molybdate, 380 cc. of concentrated sulfuric acid, and 620 cc. of distilled water. The stannous chloride solution was made up from 2.5 grams of $SnCl_2 \cdot 2H_2O$, 30 cc. of concentrated sulfuric acid, 70 cc. of distilled water, and some metallic tin. To 100 cc. samples of water and standards alike were added 1 cc. of the molybdate reagent and one drop of the stannous chloride, except that two drops of the latter solution were used when phosphate concentrations were high.

The range of salinities encountered even at a single station renders the use of salt water standards impractical so that it has been necessary to employ standard solutions prepared in distilled water. It seems generally agreed that the use of distilled water standards in estimating phosphates in water of oceanic salinities yields results which are low by approximately 25 per cent (Robinson and Wirth, 1935; Igelsrud, Robinson, and Thompson, 1936; Ibañez, 1933; and Cooper, 1938a). Cooper (1938a) has recently presented evidence to indicate that the usual factor of about 1.25 is reduced to about 1.12 when the amount of molybdate reagent is reduced from 2 cc. to 1 cc. Since the salinities in the Chesapeake Bay are on the average about half those of sea water, the use of 1 cc. of molybdate imposes a rather negligible error, at least this correction is considerably less significant than other errors inherent in the method. Standard solutions of two strengths have been found adequate to cover the range of phosphate concentrations in the Bay except for unusually high values. One of these standards contains 7.03 and the second 0.703 mg. atoms P/m^3.

Comparison is made in an enclosed frame holding the 200 mm. tubes and fitted with a colorimetric eyepiece. The use of a split-field or comparison eyepiece makes possible a much more accurate estimation and it is not unusual to obtain repeated checks to the third sig-

nificant figure when the phosphate concentration is as low as 0.0175 mg. atoms P/m³. It is important, however, to use a uniform height of column of the unknown (in our case 150 mm.) and to calibrate the instrument for different heights of the standard column, particularly when the concentrations lie below about 0.35 mg. atoms P/m³. With our instrument, for example, a standard column of 63.5 mm. would seem to indicate a phosphate concentration of 0.298 mg. atoms P/m³ whereas it represents an actual concentration of only 0.075 mg. atoms P/m³. With short comparison tubes and a column height of only 150 mm., this discrepancy is an advantage because it makes possible a more accurate estimation in the lower values and the appropriate correction is easily applied from a prepared table or graph.

The magnitude of this correction imposed by an apparent failure of Beer's Law in the lower concentrations is so great that it may affect materially the interpretations of minimum phosphate values especially in surface waters. In the Chesapeake Bay, although the amount of phosphate in surface water at any season of the year may in the course of the day become critically reduced, a measurable quantity has always been present. On occasions the quantity is so small that no corrected or "true" values can be assigned to it, but the uncorrected values are readily determinable and have never been less than 0.11 mg. atoms P/m³. That this represents the merest trace is evident from the lowest corrected value considered in the present paper which is 0.01 mg. atoms P/m³ and equivalent to 0.230 mg. atoms P/m³ uncorrected. In the Chesapeake Bay differences in concentrations below about 0.15 mg. atoms P/m³ apparently have no pronounced practical significance, but the absence of zero concentrations is interesting from a theoretical standpoint. The waters of the Chesapeake Bay like those of the oceans are probably being constantly enriched in phosphates by the activity and decay of organisms and it would appear reasonable that some quantity of such material should be present at all times. That zero concentrations of phosphate are commonly found in ocean waters may indicate, perhaps, a different balance between the component populations of organisms than occurs in the Bay or it may be due to differences in the method of estimating phosphates.

Phosphate samples, because of their included plankton, are very unstable. When exposed to the sun in clear glass for only 15 minutes the phosphate content has been noted to drop from 0.413 to 0.044 mg. atoms P/m³. When stored in dark glass bottles and protected from light the phosphate content characteristically rises above the original value for a period of 24 to 48 hours after which it falls gradually to

TABLE I

CHANGES IN PHOSPHATE CONTENT (mg. atoms P/m³) OF WATER SAMPLES STORED IN SEPARATE DARK-GREEN-GLASS BOTTLES IN DARKNESS AND SHOWING EFFECTS IN UNPRESERVED AND PRESERVED SAMPLES. RESULTS FROM ONE OF SEVERAL EXPERIMENTS, EACH VALUE REPRESENTING AN AVERAGE FROM TWO SAMPLES

November 1938	Time	No Preservative	CS_2 Preservative	NaF Preservative
6	3:30 P.M.	0.44	0.48	0.12
7	2:15 P.M.	0.54	0.49	Trace
8	1:20 P.M.	0.42	0.54	Trace
9	10:10 A.M.	0.27	0.50	Trace
10	10:00 A.M.	0.22	0.41	Trace
11	9:20 A.M.	0.24	0.43	Trace

values between about 25 to 50 per cent of the original (Table I). The desirability of analysing samples promptly after collection is therefore apparent. Several possible preservatives to reduce enzyme action in the samples have been suggested in the literature, but none has been helpful in our case. The use of 1 to 3 per cent sodium fluoride as a preservative of Bay water gives results which are immediately low by about 40 per cent and the sample rapidly changes upon standing yielding near-zero values within about 24 hours. Best results have been obtained by the addition of 3 or 4 cc. of carbon disulfide per 150 cc. of sample water, but the values, although more uniform over continued periods of time, tend to be slightly high rather than low as in the case of other preservatives (Table I).

REGION STUDIED

Water samples were collected from over the greater part of the Bay, the southermost station being located a short distance outside Cape Henry, whereas the northermost station included here is off Sandy Point below Baltimore Harbor. Data from these water samples are supported by records from additional stations selected for local characteristics in the Bay and from other stations in the Patuxent River. The location of principal Bay stations is indicated in Fig. 1, the river stations (Fig. 2) being identical with those studied by Newcombe and Horne (1938), and the usual sampling depths being 1, 9, 15, 23 meters, and bottom (50 meters or less).

The general hydrographical characteristics of the Chesapeake Bay, particularly in regard to temperatures, salinities, and plankton, as these vary with depth, time, and place, have been described by Cowles (1930); Wells, Bailey, and Henderson (1929) and by Newcombe, Horne, and Shepherd (1939). A résumé of the more outstanding aspects of these

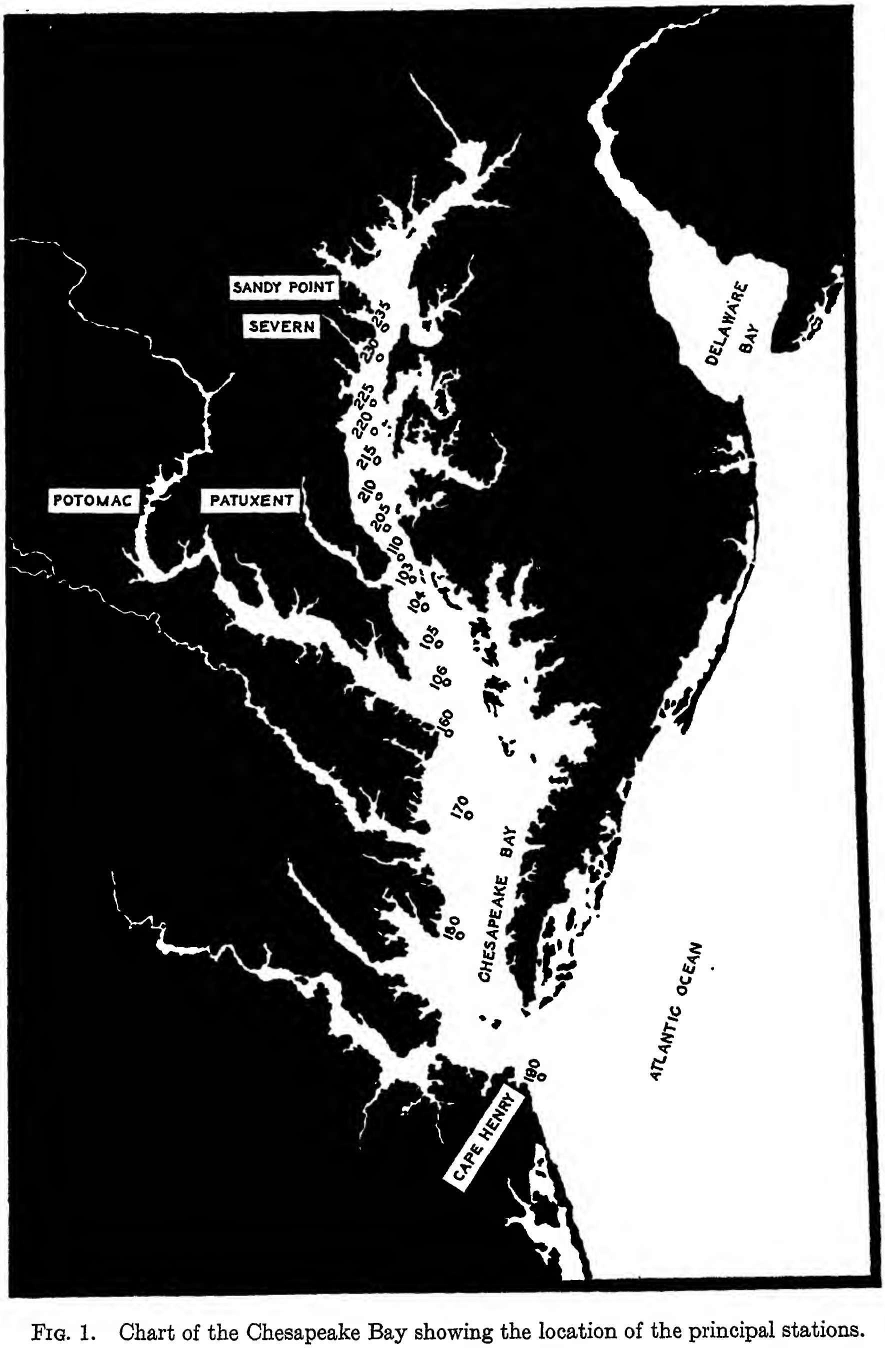


FIG. 1. Chart of the Chesapeake Bay showing the location of the principal stations.

characteristics that exist in the Bay waters, since they differ markedly from those found in the sea, will aid materially in establishing a basis for evaluating the phosphate data to be presented in the following pages. This résumé is drawn in part from the existing literature cited above and from original data obtained in the course of the present investigation.

Hydrographically considered, the Bay waters approach oceanic conditions at the southern end and fresh water characteristics at the northern end. Its tributaries both large and small, being also subjected to the action of the tides (Haight, Finnegan, and Anderson, 1930), reproduce over shorter distances the characteristics and differences encountered within the Bay proper.

Salinity. In general, the lowest salinities of the year obtain at all depths throughout the Bay during the spring period (Table II). In the middle portion of the Bay above the Potomac River the surface salinities during March are about 10.00 per mille whereas at the bottom they are about 18.00 per mille. Beginning in April the salinities gradually rise and attain maximum values (about 17.00 per mille) at the surface in November and December while the bottom salinities tend to be highest (about 28.00 per mille) in the late summer. The salinity of the Bay water is maintained, in part, by the inflow of ocean waters at the bottom and evaporation at the surface. The mean annual inflow of fresh water into the Bay in second-feet is about 88,000 and the evaporation has been estimated to be between 60 and 100 per cent so that the net contribution of fresh water occasions no marked flow out of the Bay. The movement of fresh water over the surface is greatest during the spring and summer and almost half (41,000 second-feet) of this water is contributed by the Susquehanna River at the extreme northern end of the Bay. Similarly the movement of ocean or coastal waters up the Bay below the surface is most conspicuous during the spring and early summer (Wells, Bailey, and Henderson, 1929).

The salinities also vary greatly from north to south in the Bay. Bottom salinities between 26.00 per mille and 33.00 per mille are apparently normal in the vicinity of Cape Henry, whereas the salt content varies between about 5.00 per mille and 17.00 per mille in the region of Sandy Point, and entirely fresh water may be found in the extreme upper reaches of the Bay. Although a pronounced vertical stratification with respect to salinity ordinarily prevails throughout the Bay, instances of mixing have been found in the various seasons which markedly disturb or destroy the vertical gradient, and small bodies of atypically fresher water containing unusual amounts of dissolved oxy-

TABLE II

THE VARIATIONS OF TEMPERATURE, SALINITY, DISSOLVED OXYGEN, AND INORGANIC PHOSPHATE WITH THE SEASONS AND WITH DEPTH AT STATION 103

These data are supported by similar records from other stations in the Bay.

	Meters	4-28-'38	5-11-'38	5-18-'38	5-31-'38	6-1-'38	6-16-'38	6-23-'38	6-24-'38	7-6-'38	7-25-'38	8-4-'38	8-10-'38	8-19-'38	8-25-'38	9-6-'38	9-13-'38	10-8-'38	11-30-'38	1-7-'39	3-22-'39
Temperature ° C.	1	14.8	17.4	16.8	18.7	18.0	22.9	23.4	23.2	23.6	25.7	26.8	27.8	27.2	25.7	24.4	23.2	17.7	8.8	5.6	5.6
	9	—	—	16.2	18.0	18.0	21.6	21.7	21.3	23.2	25.7	26.5	27.2	27.3	25.7	24.5	23.0	17.7	9.1	5.6	5.6
	15	—	—	—	18.0	—	20.3	—	—	22.8	24.8	25.2	24.9	26.2	24.8	24.7	24.2	17.8	—	—	—
	23	—	—	—	17.9	—	19.4	—	—	22.7	—	—	—	25.3	24.2	—	24.0	18.0	—	—	—
	50	13.5	14.8	17.0	17.1	17.2	19.2	19.6	19.7	21.5	22.6	23.2	24.0	24.3	23.3	23.4	23.9	18.4	11.8	5.9	5.6
Salinity per mille	1	10.42	11.00	13.15	12.93	12.50	13.00	12.74	12.30	13.81	14.50	14.40	14.35	14.03	14.01	15.42	15.78	17.13	18.27	15.89	9.54
	9	—	—	13.30	13.25	13.10	13.39	14.82	16.44	14.57	14.50	14.57	15.13	15.64	15.99	15.76	15.96	17.80	18.35	18.99	13.91
	15	—	—	—	17.15	—	17.43	—	—	18.36	15.19	16.30	17.55	20.40	22.66	22.51	21.66	19.20	—	—	—
	23	—	—	—	18.75	—	19.39	—	—	19.40	—	—	—	—	24.70	—	22.64	20.27	—	—	—
	50	18.47	20.55	21.48	19.66	19.50	22.93	18.56	22.40	21.65	20.97	20.16	19.44	24.29	14.70	23.69	23.69	22.10	20.65	22.15	17.24
Dissolved oxygen cc./liter	1	7.68	6.27	6.23	6.24	6.49	2.33	2.14	1.92	5.69	5.19	5.42	4.91	4.75	4.86	5.05	5.24	5.99	7.01	7.59	8.24
	9	—	—	5.65	5.76	6.69	1.83	0.76	0.31	4.83	4.92	4.88	4.22	3.47	2.38	4.38	3.66	5.67	6.89	8.12	7.00
	15	—	—	—	2.61	—	0.73	—	—	2.21	1.98	1.02	0.95	1.58	1.02	0.16	—	6.29	—	—	—
	23	—	—	—	2.33	—	0.73	—	—	1.66	—	—	—	—	1.36	—	0.50	5.00	—	—	—
	50	3.97	2.99	1.08	0.30	0.60	0.50	0.40	0.49	0.33	0.48	0.00	0.41	1.57	4.44	0.31	0.10	1.09	5.98	6.83	6.42
Phosphorus mg. atoms P/m^3	1	—	—	—	—	—	—	—	—	—	0.01	0.08	0.55	0.09	0.46	0.01	0.43	0.24	0.02	0.01	0.43
	9	—	—	—	—	—	—	—	—	—	Trace	0.49	1.49	0.10	1.93	0.26	0.33	0.04	0.04	Trace	0.10
	15	—	—	—	—	—	—	—	—	—	Trace	0.39	2.12	1.20	1.75	0.78	0.63	0.18	—	—	—
	23	—	—	—	—	—	—	—	—	—	—	—	—	—	2.30	—	0.81	0.30	—	—	—
	50	—	—	—	—	—	—	—	—	—	0.41	2.44	4.54	1.30	0.45	0.83	1.32	0.28	0.17	0.03	0.06

gen and nutrient salts may be encountered at any depth (Table II). The variation of salinities through the seasons, the vertical stratification, and the effect of mixing at Station 103 is indicated in Table II.

Dissolved Oxygen. Perhaps a wholly unique hydrographical characteristic of the Bay water and its most outstanding feature is the amount of dissolved oxygen present at the various depths through the seasons (Newcombe and Horne, 1938, and Table II). The Bay water is typically richest in dissolved oxygen (8.00 to 9.00 cc./liter) at the surface during February, March, and the early part of April, while the lowest quantities of the year (2.00 to about 4.50 cc./liter) prevail through July and early August. At sub-surface depths the highest concentrations (about 8.00 to 10.50 cc./liter) occur during January and through most of February. These amounts become gradually diminished beginning at about early March, and the lowest quantities of the year (0.00 to about 0.50 cc./liter) typify the latter part of July and the month of August in water 30 to 50 meters in depth. A relatively rapid recovery from this condition of oxygen depletion is effected during the latter part of September when about 9.00 cc./liter are found at the surface and approximately 7.00 to 8.00 cc./liter prevail to the greatest depths. Although almost complete saturation with respect to dissolved oxygen exists from late September until late February the actual amounts of oxygen at all depths increases continuously during this period partly as a consequence of temperature changes (about 17.0° C. in late September and about 1.0 to 5.0° C. in January and February). The most pronounced vertical stratification with respect to oxygen like that of salinity ordinarily exists during July, August, and early September, and is similarly subject to complete or partial destruction as a consequence of mixing. It is a matter of significance that stratification develops during that period (April to July) when the heaviest rains occur and when rains are most frequent (Cowles, 1930).

Temperature. The temperature changes through the seasons rather closely parallel those noted in regard to salinity and dissolved oxygen (Table II). The lowest temperatures (about 1.0 to 5.0° C.) generally occur during late January and early February and the highest temperatures (about 26.0 to 29.0° C.) exist during the month of August. During the spring and summer the surface waters are characteristically warmer than those at the bottom, the difference ordinarily being less than 4° C., whereas during the fall and winter the surface temperatures are normally lower than those of the bottom and the difference is a matter of about 1.0 to 2.0 degrees (compare Newcombe, Horne, and Shepherd (1939).

Since the stratification with respect to individual features such as temperature, salinity, dissolved oxygen, inorganic phosphates, and silicates, is presumably developed by different individual mechanisms, although the ultimate causes may be the same, it is to be expected that recovery of a state of stratification in these features after a period of mixing will proceed with differing rapidity. For this reason, sampling at different periods after mixing may on occasions give evidence of stratification in one characteristic and not in another and such evidence in itself aids in understanding the relative rate at which various processes take place. Evidence to be developed in subsequent pages of this paper indicate in part that salinity stratification in the Chesapeake waters reforms less rapidly than stratification of dissolved oxygen and phosphate, while temperature characteristics in this regard are less well understood (Table II).

TABLE III

AVERAGE CONCENTRATIONS OF PHOSPHORUS IN MG. ATOMS P/m³ FOR STATIONS 102–106 AND 110 IN CHESAPEAKE BAY

Meters	7–13 1938	7–25 1938	7–27 1938	8–4 1938	8–10 1938	8–19 1938	8–25 1938	9–6 1938	9–13 1938	10–8 1938	11–30 1938	1–7 1939	2–1 1939
1	0.20	0.17	0.14	0.36	0.60	0.38	1.02	0.30	0.43	0.25	0.28	0.21	0.46
9	0.26	0.17	0.11	0.42	1.07	0.40	1.47	0.36	0.77	0.27	0.29	0.21	—
15	0.47	0.22	0.16	0.54	2.09	1.61	2.86	1.45	1.26	0.33	0.33	0.19	—
23	—	—	—	—	—	—	2.74	—	1.40	0.45	—	—	—
30–50	—	0.56	—	2.48	3.31	1.80	2.87	1.81	1.77	0.43	0.33	0.26	0.48

Plankton. The plankton data included here is taken entirely from Cowles (1930), since plankton data accumulated during this study are not available in complete form at this time. The diatom populations include species which may be classed as fresh water, brackish water, and marine forms. In general, fresh water forms are not abundant, constituting only a small percentage of the total diatom population. Brackish water forms are represented in largest number by *Raphoneis amphiceros* Ehr. found principally below the Potomac, by *Nitzschia sigma* (Kutz) W. Sm., occurring widely over the Bay but in small numbers, and by *Nitzschia plana* W. Sm. and *Pleurosigma balticum* W. Sm. with a still more limited distribution. The bottom and semi-bottom or tychopelagic diatoms are abundantly represented in the Chesapeake Bay. *Skeletonema costatum* (Greve) is very abundant and widely distributed in the Bay and is little affected by salinity. Its behavior in the Bay was not that of a bottom form to which group it is most closely related, since it occurred in greatest numbers (about 26,000 per liter) at the surface in the upper portion of the Bay and in greatest

numbers (about 11,000 per liter) at 27 meters near Cape Henry. Other neritic diatoms show a similar distribution. Truly tychopelagic forms, the most important of which is *Paralia sulcata* (Ehr.), ordinarily are more abundant with increasing depth. It can be seen from Cowles' data that, although this was the case through most of the year, during September greatest numbers (1,000 per liter) occurred at the surface and fewest (none) at the bottom. No data were available for the month of August. From Cowles' data it may also be seen that all diatom forms for which information is given occurred either in smallest numbers during July and August or when present were most abundant at or near the surface level. In general, diatom forms were most abundant in the vicinity of the Potomac River and showed a conspicuous spring maximum (about 550,000 per liter) during March or April and a less conspicuous maximum during December (about 30,000 to 500,000 per liter), the lowest values (about 6,000 to 10,000), in general, occurring during the months of June to October.

Various species of *Difflugia* were found to be present throughout the year and were widely distributed over the Bay. They prevailed in largest numbers during the period from about June to October, according to the species, and were least common through the winter. The largest numbers were characteristically found at or near the surface. These protozoans usually show an increase in numbers immediately following the spring diatom maximum and continue to become more numerous as the diatoms decrease in number through the summer period.

Supplementary Data. The water of the Chesapeake Bay is not uniformly transparent throughout the year. Shortly after the spring rains begin the water becomes increasingly opaque and light penetration is poorest during the late spring and through much of the summer. At a somewhat variable time in the late summer or fall usually in September or October the water becomes rapidly cleared with relatively good visibility to considerable depths. Secchi disc readings through the period of a year, although they are subject to a variety of errors, show in a general way how the transparency of the surface water changes through the seasons (Table IV). That this variable transparency is due in part to soil sediment carried in by the rivers is readily evident from direct observation alone; it is borne out by the settling of sediment in the sampling bottles; and from the yellowish cast of the water characteristic of the streams to a much more pronounced degree after heavy rains. Since the drainage basin of rivers feeding the Bay is largely an agricultural region and consists mostly of the finely divided Piedmont

TABLE IV

SECCHI DISC READINGS TAKEN NEAR STATION 103

Date	Time	Secchi Disc Meters
6-29-'37	10:40 A.M.	1.9
7-22-'37	8:40 A.M.	2.0
7-29-'37	9:45 A.M.	2.4
8-19-'37	3:15 P.M.	2.1
9-25-'37	9:30 A.M.	2.3
12- 4-'37	11:00 A.M.	3.4
1- 6-'38	10:00 A.M.	3.6
2- 5-'38	10:40 A.M.	3.3
4-29-'38	10:00 A.M.	2.6

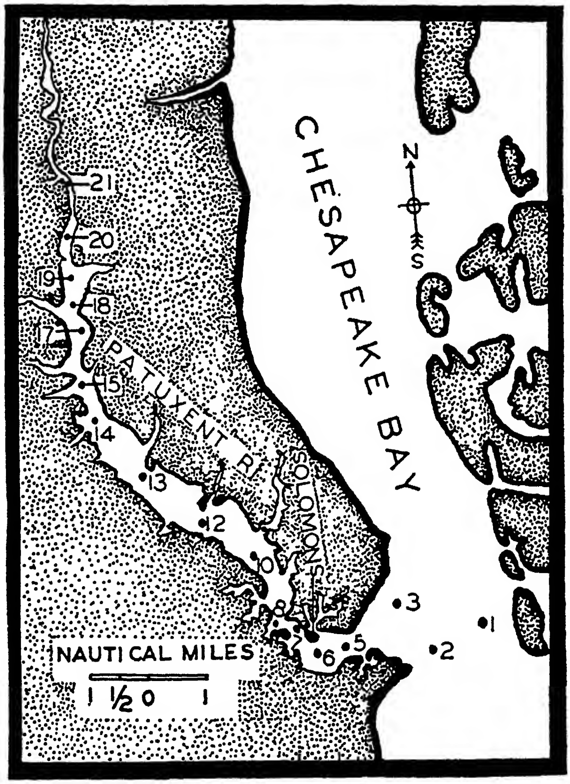
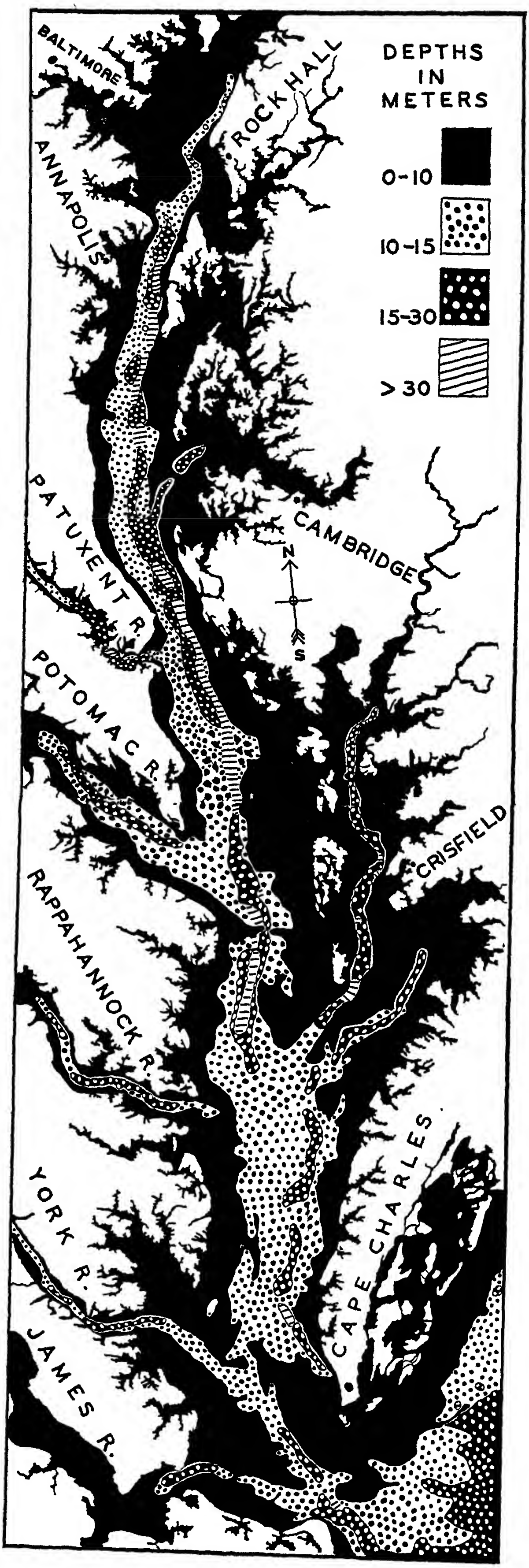


FIG. 2. Locations of principal sampling stations (numbered points) in the Patuxent River.

clays and Pleistocene flood plain deposits, the suspended matter settles out rather slowly.

RESULTS

Vertical Stratification of Phosphate. Although the Chesapeake Bay is a comparatively shallow body of water rarely exceeding 50 meters in depth (Fig. 3), during hours of daylight there exists characteristically


DEPTHS IN METERS
0-10
10-15
15-30
> 30
BALTIMORE
ROCK HALL
ANNAPOLIS
PATUXENT R.
CAMBRIDGE
N
S
POTOMAC R.
CRISFIELD
RAPPAHANNOCK R.
CAPE CHARLES
YORK R.
CAPE CHARLES
JAMES R.

through the seasons a definite and often very pronounced vertical stratification of phosphates (Fig. 4, Table II). This stratification is most severe during July, August and early September when surface waters generally contain mere traces (<0.01 mg. atoms P/m³) to about 0.5 mg. atoms P/m³ and waters at sub-surface levels, particularly between 30 and 50 meters, contain 2.0 to 4.0 mg. atoms P/m³. From late September until March, the surface concentrations are characteristically mere traces, up to about 0.20 mg. atoms, while the bottom concentrations are respectively about 0.01 to 0.50 mg. atoms P/m³. The lowest sub-surface values normally obtain during January and February whereas

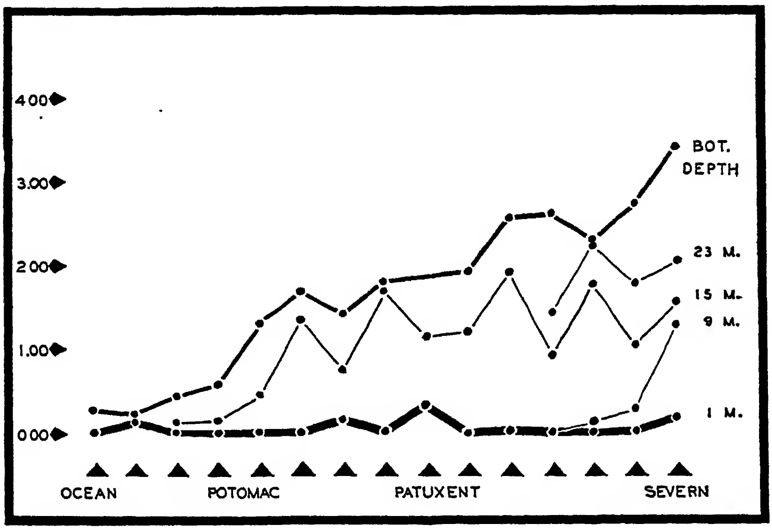


Fig. 4. The amount of phosphate as mg. atoms P/m³ present from the lower to the upper portion of the Chesapeake Bay. August 19 to 25, 1938.

the lowest surface values occur through late February and the month of March. Very low concentrations of phosphate may be present ephemerally even during the summer period, but high concentrations have not been noted in the Bay proper during the fall, winter, and early spring. The amounts present at depths of 9 meters were not infrequently less than those at the surface during the summer period whereas through other seasons lesser amounts normally occurred down to about 15 meters (compare Cooper, 1938b).

Collections from the Patuxent River, at times, also reveal a vertical stratification of phosphate (Table V) which is less permanent due to more frequent as well as more pronounced mixing (Fig. 2).

Fig. 3. Bathymetric chart of the Chesapeake Bay (Data from Coast and Geodetic Survey map).

TABLE V

Average Concentrations of Phosphorus in mg. atoms P/m³ in the Lower and Upper Portions of the Patuxent River

Date		Stations 8–14	Stations 15–21	Total Average
7-19-'38....	Average Surface	0.12	0.49	0.30
	Average Bottom	0.36	0.60	0.48
8-15-'38...	Average Surface	0.52	—	0.52
	Average Bottom	0.89	—	0.89
8-19-'38....	Average Surface	0.97	—	0.97
	Average Bottom	0.73	—	0.73
11-9-'38....	Average Surface	0.36	0.70	0.53
	Average Bottom	0.40	0.72	0.56
3-1-'39......	Average Surface	0.07	0.17	0.12
	Average Bottom	0.02	0.29	0.16

Variation of Phosphate throughout the Bay. The amount of phosphate present in surface water is essentially uniform throughout the Bay on comparable occasions, whereas at sub-surface depths greater quantities of this substance are normally found toward the north and progressively lower concentrations obtain as more oceanic conditions are approached toward the south. This condition as it existed during the summer period is illustrated in Fig. 4 where it is shown that in the coastal water off Cape Henry the quantity of phosphate at the surface was a trace and that at the bottom (30 meters) a concentration of 0.28 mg. atoms P/m³ obtained, while at Station 230 the surface water contained 0.345 mg. atoms, and the bottom water 3.83 mg. atoms P/m³. It may also be noted that sub-surface depths at this season were increasingly rich in phosphate toward the head of the Bay so that at Station 230 the water to within 9 meters of the surface contained 1.79 mg. atoms P/m³.

During the fall and winter phosphorus in the surface water is likewise relatively uniform over the Bay and greater amounts are found at bottom depths up the Bay. On September 24, 1938, at Station 230 the quantity of phosphate present at the surface was 0.08 mg. atoms, at 9 meters 0.46 mg. atoms, at 15 meters 0.56 mg. atoms, and at 41 meters it was 0.53 mg. atoms P/m³, while at station 106 on October 8, 1938, at similar depths the values were respectively 0.09, 0.03, 0.05,

and 0.41 mg. atoms P/m³. Similarly the range from surface to bottom during February and March was between 0.01 and 0.10 mg. atoms at Station 180 and between 0.30 and 0.54 at Station 235.

In the Patuxent River at Station 21, the amount of phosphorous present during the summer period was less than 1.00 mg. atoms P/m³, whereas at Station 14 about 20 kilometers downstream approximately half this quantity was found.

Daily Variation of Phosphate. Almost daily collections for phosphate analysis have been made from the Laboratory pier which extends 230 meters into the Patuxent River near its mouth. The results of

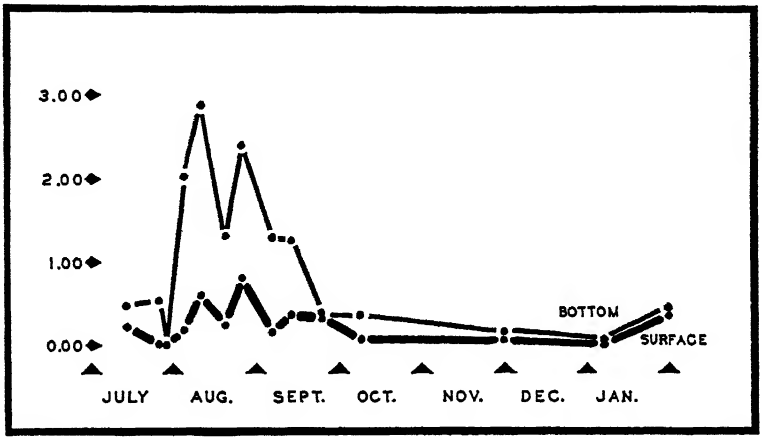


Fig. 5. The amount of phosphate as mg. atoms P/m³ present in the middle portion of the Bay during the period from July, 1938, to February, 1939. A greater seasonal difference characterized the upper portion of the Bay and a lesser seasonal variation occurred toward the south.

this sampling averaged for periods of two weeks each are shown in Fig. 6.

The daily changes indicated in Fig. 7 show that the phosphates vary

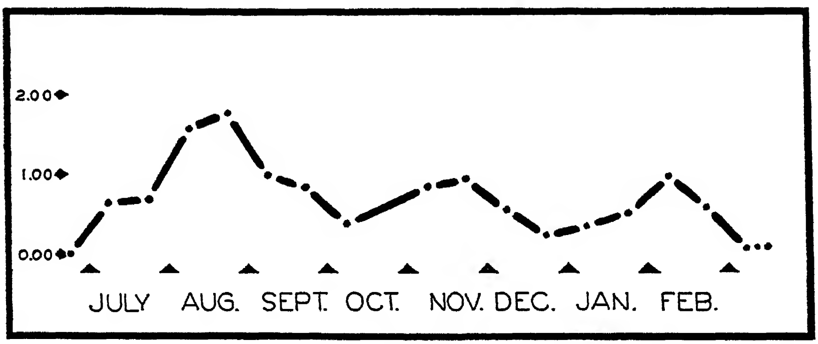


Fig. 6. The amount of phosphate as mg. atoms P/m³ present through the seasons in the Patuxent River at its mouth. The daily values averaged bi-weekly. June, 1938 to March 15, 1939.

up and down from day to day in an unpredictable fashion, the values being more uniformly low through the fall, winter, and early spring. This variation was obtained in surface water at the end of the pier where, by virtue of position, the tidal flow is slow relative to the nearby deeper waters. Since the sampling was done at about the same hour (9:00 A.M.) on each day, it is evident that tides or other types of water movement ostensibly will not account for the variation observed.

Because it has been impractical to obtain daily records from the Bay proper, the daily variation prevailing there at different depths cannot be accurately estimated. However, it would seem probable that a similar type of variation does occur in the Bay, and this is reflected in the irregularities of the curves shown in Fig. 5. That the

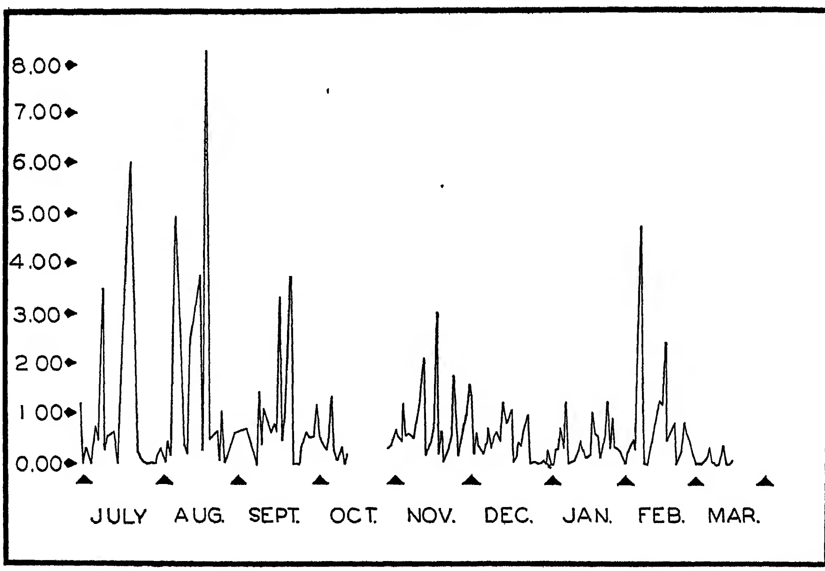


FIG. 7. The daily variation of phosphate as mg. atoms P/m^3 occurring in surface water at the mouth of the Patuxent River. June 1938 to March 15, 1939.

daily variation within the Bay is less extreme than in the river is suggested by the lesser variation shown in this Fig. and in Table II. Furthermore, on single cruises when sampling was done during most of the daylight hours the amount of phosphorus found from station to station was essentially the same for corresponding depths except as the effect of latitude was reflected in those values.

Seasonal Variation of Phosphate. Unlike the cycle of phosphorus in the sea, in the Chesapeake Bay these substances have been most abundant during the summer period (Fig. 5), a similar tendency having

been noted in the samples drawn from the pier (Fig. 7), and in the lower reaches of the Patuxent River (Table V).

In the case of the pier samples for which the daily variation is known, it may be seen that the phosphates at this point in the river increased slightly through July, attained high values (1.82 to 1.94 mg. atoms P/m^3) during August, and declined through the month of September (1.29 to 0.90 mg. atoms P/m^3) to a lower average value of 0.55 mg. atoms in the first half of October. During November a rise of appreciable proportion (1.20 mg. atoms P/m^3) occurred which, after a decline through December and early January, again occurred during early February. The lowest values of the year have prevailed through the first three weeks of March, 1939 (average 0.09 mg. atoms P/m^3). It is not known whether the fall and winter variations represent a normal response to seasonal conditions alone or whether the activities of oystermen by disturbing the bottom deposits rich in phosphorus in the immediate vicinity and for some considerable distance upstream may have been responsible for the higher values.

Collections from over the Patuxent River, although too few in number to be significant in themselves, support the seasonal trends noted at the pier (Table V). The river water showed an increase of phosphate from June through August with highest values occurring during the latter month. In November lesser amounts were found in the lower portion of the river than during the summer period but upstream concentrations remained much the same. A river collection on March 1, 1939, indicated that the phosphate content even upstream was low. This table seems to suggest that the phosphorus carried into the river by soil run-off is largely utilized within the river itself since the largest amounts (about 1.00 mg. atoms P/m^3) found in the river water in the upper reaches of brackish water at Station 21 are apparently consumed before the river water arrives in the region of Bay conditions near its mouth. Below Station 14, as an example, the water normally has been found to contain about half as much phosphate as in the region between Stations 14 and 21.

The seasonal trends within the Bay proper are indicated in Fig. 5, Table II and Table III. The values averaged in this graph are taken from the middle portion of the Bay between Stations 106 and 110, larger bottom values being characteristic through the summer in the upper part of the Bay and lower bottom values obtaining below the Potomac as is indicated in Fig. 4. An analysis of Fig. 5 shows that unusually high concentrations of phosphorus (about 2.0 to 4.0 mg. atoms P/m^3) characterized the sub-surface depths during the month of

August and early September and that, in general, high values obtained through much of July as well. The low values obtained during late July, in view of the daily variation noted at the pier, presumably represent a low point in the Bay variation whereas the remainder of the summer records for the Bay on the same basis apparently represent higher levels in such a variation. The true amounts of phosphate present therefore would be known only by the averaging of daily records as was done at the pier. The records obtained in the course of the summer would, however, appear to be sufficient to establish a reasonably accurate understanding of the amounts characteristically present. In the month of August, 2 to 4 mg. atoms of phosphorus were present not only at the bottom but often to within 9 meters of the surface. Through early September a decline in the amount of phosphate present began which first resulted in an impoverishment of the upper 10 to 15 meters of water. The decrease by late September was well advanced since the water below the 30-meter level contained on the average about 0.34 mg. atoms P/m^3. This represents approximately a ten-fold decrease in a period of about one month. From November to January only scant amounts of phosphate (about 0.01 to 0.15 mg. atoms P/m^3) were found. A collection made in February showed a slight rise (0.46 to 0.48 mg. atoms P/m^3) over the preceding month, but small amounts (average 0.14 and maximum 0.70 mg. atoms P/m^3) were found during March, 1939. On frequent occasions through the winter the amount was slightly less at depths of 10 to 15 meters than at the surface (compare Cooper, 1938b).

Hourly Variation of Phosphate. The hourly variation of phosphate was determined for a period of 25 hours at Station 110 on August 8–9, 1938. Station 110 was chosen for this purpose because of the established stations it was representative of the seasonal changes with respect to dissolved oxygen, salinity, temperature, and phosphates taking place up to that time and it was also favorably located at a point approximately in the middle of the Bay both north to south and east to west. The weather was essentially the same both days, the maximum temperatures being respectively 33° C. and 34° C. and the minimum temperatures being 24° C. and 23° C. Both days were exceptionally clear except that cumulo-nimbus clouds intermittently obscured the sun between 3:00 and 5:15 P.M. on the 8th. A brisk northwest wind blew steadily during the 8th until about 10:00 P.M. after which moderate calm prevailed.

The amount of phosphate present in the water varied greatly during the 25-hour period (Fig. 8). The surface values during the day of the

8th showed a steady decrease from 4:00 A.M., when 0.43 mg. atoms P/m³ were present, to 0.10 mg. atoms at 3:00 P.M. with only minor fluctuations up to this time. Between 3:00 and 5:00 P.M. an increase was noted to 0.41 mg. atoms and at 6:00 P.M. the concentrations were again low (0.13 mg. atoms P/m³). From 6:00 P.M. and until 10:00 P.M. the amounts of phosphate increased very greatly, as much as 3.30 mg. atoms being found at the latter hour. Since sunset occurred at approximately 7:15 P.M. and darkness prevailed from about 8:00 P.M., this increase developed in association with the first hours of night. Between 11:00 P.M. and 2:00 A.M. a considerable decrease in surface phosphate occurred down to 0.45 mg. atoms P/m³ which at 5:00 A.M. had increased slightly to 0.59 mg. atoms P/m³. The amounts present

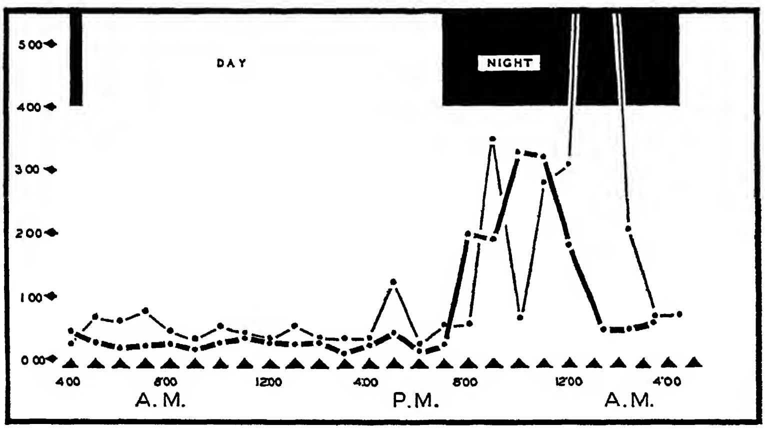


FIG. 8. The hourly variation of phosphate as mg. atoms P/m³ occurring at Station 110, August 8–9, 1938. The surface variation is indicated by the heavy black line and the bottom values at 16 meters are shown by the lighter black line.

between 2:00 A.M. and 5:00 A.M. were therefore approximately the same as those found at 4:00 A.M. on the preceding day.

The amount of phosphate at the bottom depth of 16 meters varied similarly. During the hours of daylight the amounts present became reduced from about 0.70 mg. atoms in the morning to 0.33 mg. atoms at 3:00 P.M. At 5:00 P.M. 1.23 mg. atoms prevailed, at 6:00 P.M., 0.22 mg. atoms were found, and thereafter a conspicuous increase occurred attaining a maximum value of 12.00 mg. atoms P/m³ at 2:00 A.M. At 3:00 A.M. 2.10 mg. atoms were still present but at 4:00 A.M. and shortly before sunrise (about 4:20 A.M.) 0.68 mg. atoms P/m³ were found. The amounts at 4:00 A.M. and 5:00 A.M. (0.70 mg. atoms P/m³) were essentially those of the early hours on the preceding day. The concentration

present at 10:00 P.M. was lower (0.65 mg. atoms P/m^3) than during the other hours of darkness.

The quantities of phosphate occurring at an intermediate depth of 8 meters closely approximated the concentrations found at the surface from 4:00 A.M. to 9:00 P.M., but between 10:00 P.M. and 3:00 A.M. they varied up and down between 3.20 and 0.43 mg. atoms P/m^3. At 4:00 and 5:00 A.M., however, the amount of phosphate at this depth was about 0.60 mg. atoms P/m^3, being similar to the surface and bottom concentrations.

The temperature variation at the surface over the 25-hour period was between 26.8° C. and 27.4° C. At 8 meters it was between 25.9° C. and 27.6° C. and at the bottom it was between 24.8° C. and 26.2° C. No definite temperature or salinity trends occurred. The salinity variation at the surface was 14.00 to 14.43 per mille, at 8 meters it was 13.82 to 14.42 per mille, and at 16 meters it was 14.32 to 16.91 per mille. The changes in pH as determined with a glass electrode at the surface during the day were between 7.64 and 8.00, whereas throughout the night a severe drift from less than pH 4.50 to about 7.80 indicated the presence of large quantities of CO_2 in the water, thus testifying to the magnitude of the vital processes in operation. The pH values during the day at 8 meters were about those at the surface and at the bottom they were around 7.20. During the night the pH drift at these depths was less than at the surface, being between about 6.80 to 7.30. The pH obtaining at various depths cannot be precisely determined until a method is devised to avoid exposure of the sample to air. Colorimetric methods are obviously inadequate.

Since the significance of the hourly variations in phosphate content noted above was not fully appreciated during the summer no other records of this character were obtained for that period. In order to determine whether similar changes occurred within the period of diatom maximum and zooplankton minimum (particularly protozoans), samples were collected from the laboratory pier at one-hour intervals from the late hours of darkness (4:00 A.M.) to the earlier hours of daylight (9:00 A.M.) and also from the late hours of daylight (4:00 P.M.) to the early hours of night (10:00 P.M.) on three successive days, April 9, 10, 11, 1939. On April 9, 1939, the amounts of phosphate in the surface water at the various hours were as follows: 4:00 A.M., 0.95 mg. atoms; 5:00 A.M., 0.46 mg. atoms; 6:00 A.M., 0.20 mg. atoms; 7:00 A.M., 0.14 mg. atoms; 8:00 A.M., 0.02 mg. atoms; and 9:00 A.M., 0.015 mg. atoms P/m_3. Sunrise was at 5:35 A.M. but the first sunshine was between 6:00 and 7:00 A.M. after which the sky was entirely clear. The water tempera-

ture was between 8.3 and 8.5° C. That evening the phosphate concentrations were as follows: 4:00 P.M., 0.018 mg. atoms; 5:00 P.M., 0.01 mg. atoms; 6:00 P.M., 0.09 mg. atoms; 7:00 P.M., 0.03 mg. atoms; 8:00 P.M., 0.01 mg. atoms; 9:00 P.M., 0.10 mg. atoms; and 10:00 P.M., 0.432 mg. atoms. Sunset was at 6:37 P.M. and the water temperature was between 8.8° C. and 9.8° C. Identical sampling on the two succeeding days when the water temperature gradually increased to between 10.0° C. and 12.0° C. showed, likewise, that a decrease in the amount of phosphate occurred as the light from the sky increased and that larger amounts of phosphate were present after the first few hours of darkness. The variations in phosphate present between the hours of light and darkness in this spring period, although they showed the same tendency noted during the summer, were variations between lower extremes. At 7:00 P.M. on April 11, 1939, the amount of phosphate present in the water was 0.134 mg. atoms P/m^3, in harmony with the amounts found earlier and later. A single ctenophore about 15 mm. in size in one of the tubes used in the estimation of phosphate at this hour was accompanied by an increase in phosphate (0.37 mg. atoms) in that tube, about three times that found in the check sample.

Discussion

It is known that dissolved phosphates in ponds, lakes, rivers, and in the sea are utilized by phytoplankton in the presence of sunlight. Particularly in the surface strata of oceans, where a greater stability of conditions exists than elsewhere, the amount of phosphate present is greatest during the winter period of poor illumination and least during the summer when illumination is strongest, and the phytoplankton varies accordingly. In the Chesapeake Bay dissolved inorganic phosphates are most abundant during the summer and least abundant through the fall and winter, attaining perhaps their minimum values at the surface levels in the early spring. That the seasonal cycle in regard to phosphates in the Chesapeake Bay is the reverse of that found in the sea can perhaps be explained best on the basis of light availability to the phytoplankton in the Bay.

From the records of hourly sampling it is seen that light and darkness apparently produce distinct effects upon the amount of phosphorus present in the water. In the case of the 25-hour sampling, only moderate or small amounts of phosphate were found and apparently a normal utilization throughout the day gradually reduced the quantities present. Between 3:00 and 4:00 P.M. when clouds obscured the sun, an increase in phosphate was noted at all depths. With the passing

of the clouds the quantities of phosphate again became reduced apparently in response to better illumination. Directly with the setting of the sun the amounts of phosphate increased greatly and large amounts were prevalent in the sub-surface water until shortly before sunrise and small amounts obtained in the early hours of daylight on the second day. In the surface water the increase in the amount of phosphate present also began with the setting of the sun but the decrease began earlier than at the bottom. These changes are distinctly associated with periods of light and darkness and cannot be explained on a purely physical basis of water movement. It is significant that this cycle of change occurred at a point in the Bay protected from the influence of city pollution or a rich source of phosphate supply. That changing tides fail to explain this cycle is apparent from the uniformity of values through the day and the character of the values observed through the night, there being a change in the direction of tide flow about every $6\frac{1}{2}$ hours. Furthermore, there is no conspicuous mass movement of water down the Bay as is the case in rivers, although aside from the tidal movements, the general tendency is for surface water to move down the Bay and bottom water to move toward the north—a movement made evident only over periods of weeks or even several months. That these fluctuations in phosphate content are the consequence of intrinsic processes operating in the water is further indicated by salinity characteristics. The decrease in phosphate content of surface waters after 11 P.M. could be due to the inflow of waters of different biological characteristics, and, in fact, is suggested by the change of tide from flood to ebb that took place at this hour. The changes observed in this period of 25 hours would seem to indicate that not only is there an absence of phosphate utilization during the night, but that, also, a pronounced liberation of phosphorus occurs at such times.

That phosphates are liberated in darkness by planktonic organisms in water from the Bay is borne out by experiments on the stability of samples in clear glass and in dark-green-glass bottles. Without exception the amount of phosphorus has increased in samples contained in dark glass and this increase has persisted for 24 to 48 hours. Following this period the phosphate concentration decreased as a result of the rebalancing of the living populations which vitiated the further value of this procedure as an experiment in this connection (see Renn, 1937). On the other hand, the amount of phosphate became rapidly reduced in samples contained in clear glass and on one occasion became a mere trace within an hour after collection. The response of phosphate concentration to light and darkness in the hourly sampling during April,

although the variation involved smaller quantities than during the summer, was an even greater percentage difference than that of the summer.

That regeneration of phosphorus takes place constantly as a result of the decay of dead organisms in the sea is generally known (Seiwell, 1935), but the possibility that normal metabolic processes in phytoplankton or zooplankton may bring about such a regeneration is apparently less generally appreciated. Seiwell (1935) has adduced evidence that the average phosphorus content of plankton from the North Atlantic is about 2×10^6 times that of the sea water containing them. In the Chesapeake Bay, the maximum phosphorus content at any season is rarely more than 500 times the minimum content and more typically it is about 10 to 100 times the minimum. If similar ratios as those given by Seiwell also apply even approximately in the Chesapeake Bay, the loss of about 1/2000th of the phosphate in the plankton would be adequate to account for the day and night differences noted here and a much smaller percentage would explain the day and night differences observed during April. Although the regeneration of phosphorus during the hours of darkness noted during the summer may be due to the decay of dead organisms alone, that so rapid a liberation of phosphorus would occur by this process does not appear likely in view of the depths in the ocean where this process is known to produce noticeable effects (Seiwell, 1934; Igelsrud, Robinson, and Thompson, 1936). Furthermore, that decay alone will not adequately explain such an increase is indicated by the delayed decrease of phosphate concentration in sample bottles placed in the dark, this decrease, as shown by Renn (1937) and in part by our experiments (Table I), being associated with a increase of bacteria—and the bacteria presumably increase as the decay of normal component organisms progresses.

Renn (1937) has suggested the possibility that diatoms and other phytoplankton organisms may bring about direct regeneration. His attempts to demonstrate such regeneration gave negative results, but he was of the opinion that the possibility nevertheless exists. His failure to obtain positive results may have been due to an inappreciation of the possible rôle of light in the process since no mention of light conditions was given. In this connection, the findings of Ketchum (1938) concerning the marine diatom *Nitzschia Closterium* would appear significant (see also Harvey, 1933). He found that when cultures of this organism were grown in a medium devoid of phosphate and in the presence of light, and were subsequently transferred to a phosphate-containing medium and placed in the dark, the organisms absorbed

phosphate rapidly and in an amount proportional to the time they were exposed to light in the phosphate-free medium. It is further stated that the absorbed phosphate combines loosely with water-soluble organic compounds in the cells. These findings would tend to indicate that the utilization of phosphate by this organism is not wholly brought about by an immediate effect of illumination, but rather that the absorption of phosphate constitutes an important step within the organism in the disposal of the immediate or secondary products of photosynthesis. Whether the rôle of phosphates within the individual organism may be one of translocation, transformation, assimilation, or some other process is, however, immaterial to this discussion. But if under favorable conditions, that is, an abundance of phosphate, such an organism would rapidly absorb large quantities of phosphate and combine these loosely within the cell, it would seem possible that a mechanism exists within the diatom which could afford a rapid and easy elimination of this substance. One might further postulate that, if such a mechanism exists, under conditions where severe competition is taking place for dissolved phosphate as a consequence of large populations of diatoms, then the amounts of phosphate so loosely combined would be small or lacking and the regeneration would necessarily be less marked. Aside from these considerations, the results of Ketchum will aid in explaining the slow rise in phosphate concentration in the hourly sampling of April, 1939 with the first hours of darkness, since at this time and in view of the low amounts of phosphate found during the day in a season of high diatom numbers there was almost certainly a "phosphate debt" like that experimentally devised by Ketchum.

It remains to be demonstrated whether the phosphate regeneration observed during the summer in the Chesapeake Bay is a consequence of decay process, of regeneration by living phytoplankton, or regeneration by living zooplankton. Since in the Chesapeake Bay the number of protozoans is greatest during summer, it is possible that these organisms bring about the regeneration of phosphorus noted.

The effect of light and darkness upon the amount of phosphate present in the water of the Chesapeake Bay helps materially to explain the peculiar seasonal cycle of phosphorus found here. It is seen that the period of high phosphate concentrations coincides with the period during which light penetration is poorest (Tables II, III, and IV). It is during this period, also, that the amount of dissolved oxygen is least and the decrease of phosphate in the late summer is paralleled by an increase in dissolved oxygen. The absence of measurable quantities of oxygen and the apparent scarcity of phytoplankton at the greater depths

through August and early September testify to the failure of light to penetrate to those depths since the distribution of diatoms at the bottom as in the Bay in general cannot be explained on the basis of salinity or temperature characteristics. That light from the sky is brightest during this period is apparently more than counteracted by the reduced penetration occasioned by the turbidity of the water. During the fall, winter, and early spring, when the water is relatively more transparent, the small amounts of phosphate present even at the greatest depths are paralleled by high concentrations of dissolved oxygen indicating the penetration of light to the bottom.

That the higher concentrations of phosphate in the Bay during the summer cannot be entirely explained by a richer contribution of phosphate from the rivers is suggested by the percentage of phosphate apparently consumed within the river itself and by the highest phosphate concentrations found even at considerable distances upstream—values which have not exceeded 1.00 mg. atoms P/m^3, when the Bay water during this period contained commonly 2.00 to 4.00 mg. atoms P/m^3.

The fact that the Bay water is very rich in phosphorus in the north during the summer period and progressively poorer in this constituent toward the south might seem at first to be the result of a greater proximity to a source of phosphate supply in the north and a successively more complete absorption of this substance as the water moves toward the south. However, the movement of water is not uniform down the Bay at all depths since the fresher surface water tends to move down the Bay and the denser and more saline bottom water of coastal origin moves toward the head of the Bay—except for tidal effects (Cowles, 1930). It would seem more in agreement with the relations already described to suggest that the water from the Susquehanna River, comprising as it does about half the water brought in by all rivers, spreads out over the surface of the Bay in its upper reaches and the silt carried by it at times of heavy rains slowly filters toward the bottom in lesser and lesser degree as the water moves southward. The greater quantities of phosphate found in the upper reaches of the Bay can then be explained as being the consequence of a more restricted utilization and a greater regeneration of phosphate and that diminishing phosphorus down the Bay results from successively less regeneration and more complete absorption as light penetration improves toward the south. Because the Potomac River, with a mean annual inflow of about 15,000 second-feet, in its lower 100 kilometers is almost as broad as the Bay, it is seen that suspended matter carried from upstream would be rather

well settled out before this water enters the Bay proper. This would be consistent with the fact that greatest numbers of diatoms are characteristically found at the junction of this river and the Bay (Cowles, 1930) and would explain the sudden decrease in phosphate concentration at this point noted in Fig. 4.

Varying illumination either from the sky or as a consequence of varying transparency of the water also provides a basis for explaining the great change in phosphate concentrations that often occur from day to day in the samples collected from the Laboratory pier (Fig. 7). As is the case of the Bay proper, it would be difficult to explain on a physical basis of water movements how the river water at the pier could contain 0.45 mg. atoms on one day, 8.45 mg. atoms the second, 4.57 mg. atoms on the third, and 0.95 mg. atoms P/m^3 the fourth day. The elimination of phosphate from planktonic organisms at times of poor illumination would provide a suitable mechanism for this otherwise unpredictable variation. The high temperatures ranging from about 26° C. to 30° C. which obtain in the Bay during the summer would in addition insure unusually high rates of plankton metabolism consistent with this explanation and the much lower temperatures of about 1.0° C. to 5.0° C. occurring during the winter would apparently favor a less pronounced response.

To what degree the amounts of phosphorus present during the winter, particularly in the deeper Bay waters, may be influenced by continued periods of heavily overcast skies and low illumination in general is not known.

In conclusion, it is pointed out that a final deduction regarding the effectiveness of the several mechanisms postulated in this paper must, of necessity, await an understanding of the precise rôle played by light under the unique complex of conditions described. It may be pointed out that the poor penetration of light during the period when light from the sky is brightest implies that both the plant and animal life of the Chesapeake Bay are materially impoverished to distinctly sub-normal levels and probably this impoverishment is especially severe as a consequence of the long and apparently continuous period of reduced food manufacture prevailing through the summer. River waters are only intermittently turbid whereas those of the Bay are subject to summation effects which are known to produce conditions unfavorable for the growth of certain plankton-feeding organisms. It would seem likely that the summer conditions noted above have not always characterized the Bay and its tributaries in their present intensity, but that the clearing of land for agricultural purposes, resulting in silt-laden streams

during the spring and summer seasons of heavy rains, has accentuated the situation, imposing deleterious effects of undetermined proportion.

Summary

1. The seasonal variations of phosphate in the Chesapeake Bay are characterized by maximum values (about 2–4 mg. atoms P/m^3) occurring during the summer season and by minimum concentrations (below about 0.30 mg. atoms P/m^3) occurring during the fall, winter, and early spring.

2. Notably greater quantities of phosphate (3–4 mg. atoms P/m^3) were found at sub-surface levels in the upper portion of the Bay and, in general, progressively lesser amounts (about 3.00 to 0.30 mg. atoms P/m^3) were found at all sub-surface depths at successive stations to the south. The surface concentrations (about 1.00 mg. atoms P/m^3 to a trace) were remarkably uniform over the Bay at comparable periods.

3. Sampling at one-hour intervals during the summer season of high phosphate concentrations indicated that, in the hours of daylight, a gradual absorption of these substances occurred, reducing the amounts appreciably. Beginning with sunset and persisting until sunrise, the Bay water was unusually rich in dissolved phosphate, but promptly at sunrise the concentrations became materially reduced in bottom samples.

4. Various lines of evidence indicate that high concentrations of phosphorus found to prevail through the summer period may be due to an accelerated regeneration and a reduced utilization as a corollary of poor light penetration. High concentrations of phosphorus found in the upper portions of the Bay and progressively lesser amounts toward the south could on this basis be explained as a result of increasingly greater light penetration toward the south as suspended matter brought in by the Susquehanna River settles out during the general southward movement of the fresher surface water.

5. The method of phosphate analysis adapted to the variably saline water of the Bay is described briefly and a discussion of several sources of error is included.

Literature Cited

Cooper, L. H. N., 1938a. Salt Error in Determinations of Phosphate in the Sea Water. *Journ. Mar. Biol. Assoc.*, Vol. XXIII, No. 1, 171–177.

Cooper, L. H. N., 1938b. Phosphate in the English Channel, 1933–8, with a Comparison with Earlier Years, 1916 and 1923–32. *Journ. Mar. Biol. Assoc.*, Vol. XXIII, No. 1, 181–195.

Cowles, R. P., 1930. A Biological Study of the Off-shore Waters of Chesapeake Bay. *U. S. Dept. Commerce, Bur. Fisheries Bull.*, Vol. XLVI, No. 1091, 277–381.

HAIGHT, F. J., FINNEGAN, H. E. AND ANDERSON, G. L., 1930. Tides and Currents in Chesapeake Bay and Tributaries. *U. S. Dept. Commerce Coast and Geodatic Survey, Special Bull.* No. 162.

HARVEY, H. W., 1933. On the Rate of Diatom Growth. *Journ. Mar. Biol. Assoc.*, Vol. XIX, pp. 253–275.

IGELSRUD, I., ROBINSON, R. J. AND THOMPSON, T. G., 1936. The Distribution of Phosphates in the Sea Water of the Northeast Pacific. *Univ. Wash. Publ. Oceanogr.*, **3**, No. 1, 1–34.

KETCHUM, BOSTWICK H., 1938. The Phosphate Debt of Diatom Cultures. Paper presented before the Physiol. Sect. of the Bot. Soc. Amer., Richmond, Virginia, December 30, 1938. Abstr. in *Amer. Journ. Bot.*, **25**, No. 10, 14s.

NEWCOMBE, C. L. AND HORNE, WILLIAM A., 1938. Oxygen-poor Waters of the Chesapeake Bay. *Science*, **88**, No. 2273, 80–81.

NEWCOMBE, CURTIS L., HORNE, WILLIAM A. AND SHEPHERD, BOLAND B., 1939. The Physics and Chemistry of Estuarine Waters in Chesapeake Bay. *J. Mar. Research* (in press).

RENN, CHARLES E., 1937. Bacteria and the Phosphorus cycle in the Sea. *Biol. Bull.*, Vol. LXXII, No. 2, 190–195.

ROBINSON, REX J. AND WIRTH, HENRY E., 1935. Photometric Investigation of the Ceruleomolybdate Determination of Phosphate in Waters. *Ind. Eng. Chem.* (Anal. Ed.), Vol. VII, No. 3, pp. 147–150.

SEIWELL, H. R., 1935. The Annual Organic Production and Nutrient Phosphorus Requirement in the Tropical Western North Atlantic. *Journ. du Conseil Intern. pour L'Explor. de la Mer.*, Vol. X, No. 1, 20–32.

WELLS, R. C., BAILEY, R. K. and HENDERSON, E. P., 1929. Salinity of the Water of Chesapeake Bay. *U. S. Geological Surv., Prof. Paper* 154-C, 105–152.

FEATURES OF GROWTH-CONTROL IN TREES

D. T. MacDOUGAL

Director (ret.), Desert Laboratory, Carnegie Institution of Washington

(*Read April 20, 1939*)

Abstract

Current generalizations as to the localized origination, polar translocation and differential effects of growth-promoting substances based principally on observations on seedlings, cuttings and small plants have been tested by dendrographic measurements and special experiments with trees.

The results of defoliation and disbudding tests make it evident that auxins which may promote growth of cambium or secondary meristem and inhibit growth of primary meristem in buds may originate in buds and in leaves. The removal of terminal buds may be followed by cancellation of the inhibiting action of their products on laterals, but by undisturbed activity of cambium. Defoliation which is followed by some awakening of axillary buds has a very positive detrimental effect on wood-formation by cambium.

While gradual downward extension of activity has been found in twigs and branches, simultaneous inception of growth through the length of tall trunks renders untenable the conclusion that auxins or their activating effects are transmitted downward immediately following their origination in leaves and buds at the beginning of the season.

Actual translocation of auxins at a rate which may reach 10 mm. per hour and their accumulation in or near the entire cambium sheet of trunks is indicated. Growth promoting action may await the incidence of favorable factors such as temperature, heightened hydration or arrival of substances in solution in the sap stream. Inception of cambial activity in trunks may precede growth in buds or dependent segments of twigs in some deciduous and evergreen trees. The possibility of production of growth promoting substances in cambium is not excluded.

Such accumulation of auxins may also result from upward movement of auxins and other material, originating in root-sprouts which may activate and maintain cambium below girdled redwoods for 21 seasons.

Apical growth and cambial thickening in roots may precede similar action in shoots or trunks by weeks or even months, or may occur in alternation, making it highly improbable that growth promoting substances originating in stem terminals are concerned in the growth of roots except in the basal sections. The flaring bases of trunks and the basal regions of attached roots act as a unit in growth, enlargement being accelerated by the effects of flexion due to wind action.

Growth of primary generative cell-masses in root- and stem apices and growth in secondary meristem may proceed independently, though coincidental during the greater part of the season. Elongation in both members begins before cambial action in the majority of species, and generally comes to an end with attendant closure of buds before cambial action has ceased.

Activation of lateral buds in the extension of branches consists in the awakening of primary meristem which is controlled by the inhibiting action of auxins from the terminal buds. Laterals of roots originate from initials in the pericycle and they may appear in great number especially in the pines following the inception of growth in length.

Many features of the growth sequences in large plants such as trees at the summit of the biological curve of development are not explainable by the results of tests of the action of plants in the juvenile stage as in seedlings, or in a regenerating condition as in cuttings.

The researches of the last decade have established the presence in seeds, buds, leaves, inflorescences, cambium layers and root-tips of sub-

stances which promote cell-division and enlargement in secondary generative layers and which accelerate callus formation and differentiation on injured surfaces. A wide variety of synthetic compounds will cause similar reactions.

Whatever the place of normal origin, their formation may also be induced by injuries and modified by temperature, light, gravity and mechanical tensions in various regions.

On the basis of facts obtained by observation of the occurrence and polar transport of growth promoting substances in seedlings the theory that the activity of the entire plant is controlled by substances originating in apical regions or foliar organs, the effects of which extend not only to stem-bases but to root-tips, has been advanced. While it is customary to designate such substances as hormones, yet it is by no means clear that their effects are transmitted in accordance with such a conception.[1]

When the attempt is made to apply hypothetical schemes evolved by the study of cuttings, seedlings and small plants to the seasonal activities of old and large plants such as trees many anomalous occurrences are encountered.

Non-conformable Features of Growth

A review of the work of competent anatomists and the results of my own dendrographic measurements during the last two decades brings to light the following features not explainable by the terms of currently accepted generalizations.

a. Although activity of the primary generative cells in buds is assumed to initiate growth, confirmed instances of beginning growth in cambium preceding awakening of the buds, as illustrated in both deciduous and evergreen trees, such as *Pinus*, *Pseudotsuga*, *Picea*, *Larix*, *Robinia*, *Gymnocladus*, *Gleditsia*, etc., are known. Such precession is not invariable in any species but may occur in one season and not in another. Inception of cambial growth in the trunk at the base of the crown, or at the base of the trunk has been noted.

b. Extension of incipient growth downward from buds to a distance of a meter or two has been observed, but measurements of trees furnish no evidence that activation of the cambium is propagated downward in trunks from the bases of the branches or from a growing region at the base of the crown.

[1] F. W. Went. "Specific Factors Other than Auxin Affecting Growth and Root-formation." *Plant Physiol.*, 13, 1938, 55–80. Kenneth V. Thimann. "Hormones and the Analysis of Growth." *Plant Physiology*, 13, 1938, 437–449. P. W. Zimmerman and A. E. Hitchcock. "The Combined Effect of Light and Gravity on the Response of Plants to Growth." *Cont. Boyce-Thompson Inst.*, 9, No. 5, 1938, 455–462.

c. Determinations of the concentration of growth promoting substances in small trees show a region of high concentration in the twigs and branches, a second in the basal parts of trunks and a third in the terminal regions of roots.

d. Activity of the apical cones of primary generative cells and of the cambium proceed independently. Elongation usually ceases before the growth of the cambium, but not invariably. Inception of secondary or mid-summer elongation of oak twigs is not accompanied by any modification of the rate of cambial action, except in the immediate vicinity of the points. Citrus constitutes a special case in which elongation of the shoots in the spring is at an end before any accretion to the trunk by cambial action. A second cycle of elongation begins at the maximum of cambial activity in mid-summer. After a quiet period a third cycle of growth in length takes place after cessation of cambial action early in October.

e. Elongation of roots begins at various stages of the season and may precede bud or cambial activity in these organs and in trunks by many weeks. Periods of growth in length of roots were seen to be coincidental with dormant periods of buds of citrus.

f. Cambial action in roots and origination of laterals follows activity of the tips, and precedes or may be nearly coincidental with cambial action in trunks. Action in the basal part of the trunk and of basal parts of attached roots is unified.

g. The relative parts played by leaves and buds in the production of growth promoting substances appears to vary. The concentration of such substances in buds may be less than in adjacent regions of twigs at the time of inception of seasonal elongation.

h. The inhibitory action of products of terminal buds on the primary generative cell-masses of lateral buds, is in contrast with origination of secondary generative cells after elongation of the main tips has begun. The growth promoting substances in both cases are seen to be inhibitory to primary, and activating to secondary cell-masses.

i. Reduction of leaf-surface by any cause at the beginning of or during the season of cambial activity is usually followed by lessened wood-production. On the other hand removal of the leaves of larches (*Larix*), which are deciduous conifers, early in the season is followed by the construction of an excessively thick layer of wood.[1]

[1] Avery, Burkholder and Creighton. *Production and Distribution of Growth Hormone in Shoots of Æsculus and Malus, and its Probable Role in Stimulating Cambial Activity, Amer. Jour. of Bot.*, **24**, 51, 1937. H. Soding. "Wuchstoff und Kambiumtätigkeit der Baume." *Jahrb. f. wiss. Bot.*, **24**, Hft. 4, 1937, 639–670. W. A. Zimmermann. "Untersuchungen uber die raumliche und zeitliche verteilung des Wuchsstoffes bei Pflanzen." *Zeitschrift f. Botanik*, **30**, 1936, 211–252.

Any further appraisement of the features cited above may be best made in connection with the results of some experimental operations arranged to identify the regions in which growth promoting substances may originate, their differential effect on primary and secondary generative cell-masses, the relative parts played by buds and leaves in the control of growth and the possibilities of the upward translocation of leaf-products inclusive of nutritive material and growth promoting substances originating in root-tips or in leafy sprouts arising from roots.

No generalizations as to the origination or translocation of growth promoting substances or no theory of hormonic control of growth can be taken to be adequate or of great value which can not be applied to the long axes of large plants such as trees, or to any plant at the summit of its biological curve.

Complete Defoliation of Pines

The most extensive series included defoliation of the Monterey pine at various stages of the season, and disbudding in the winter dormant period. The buds of this tree usually awaken in January and the subsequent elongation comes to an end in midsummer. Cambial action which begins about the time of awakening of the buds may continue until September or October.

Although the awakening of the terminal buds usually precedes cambium activity yet the dendrographic records extending over 20 years includes instances of diametral increase of the basal regions of trunks while the buds were still tightly sealed.

A survey of the effects of removal of leaves of evergreen conifers brings out clearly that such results vary with the stage of the season in which defoliation is carried out. Removal of all leaves of the Monterey pine in the autumnal dormant season, at which time disbudding was without serious effect on the cambium, is invariably followed by the death of the tree in the ensuing season as has been amply demonstrated.

Removal of any considerable part of the crown of a Monterey pine over 20 years old is usually followed by its death. Haasis records a case in which removal of the upper half of the crown of a Sequoia was followed by inactivity of the cambium on the basal region of a small trunk for a decade.[1]

If removal of mature leaves is deferred until awakening of buds has begun with emergent green leaves activation of cambium occurs and wood formation follows.

[1] MacDougal, D. T. "Studies in Tree Growth by the Dendrographic Method." *Publ. No. 462, Carnegie Inst. of Wash., Tree Growth.* Leiden, 1938.

A hitherto unreported series of defoliations of the Monterey pine by Dr. F. W. Haasis included effects consonant with the above. A small tree 1.7 m. in height, with the trunk no more than 10 mm. in diameter near the base, was defoliated on Jan. 28, 1932, by the removal of 10,600 leaves. The terminal had already reached a length of 22 mm. with exposure of the developing leaves which were left untouched. The terminal segment reached a length of 78 mm. by September, a gain of 56 mm.

Another tree (No. 35), of a size and age similar to No. 34, was defoliated Feb. 27th by the removal of about 9,200 leaves of the suites of 1930 and '31. The main terminal which was 26 mm. in length on Jan. 28, was 50 mm. in length on Feb. 27. Cessation of elongation came late in May with a length of 112 mm.

Pine No. 36 was slightly smaller than Nos. 34 and 35 and was estimated to be 9 years old. While the new leaves had made a growth of between one-fourth and one-third of their mature length, the new segment of the leader was 14 mm. long Jan. 28, and 85 mm. long when 7,700 old leaves were removed on March 30. A length of 100 mm. was reached late in May.

No. 37 was about 13 years old and slightly larger than the above. The leader was 30 mm. in length on Jan. 28. When defoliation was carried out by the removal of 9,300 old leaves on May 5, a length of 138 mm. had been attained. The new leaves were approaching maturity. No further growth in length took place.

No. 38, 2.1 m. in height and 10 mm. in diameter, standing near No. 37, was used for comparison. The new leader was 19 mm. long Jan. 28, and reached a length of 150 mm. by the end of May, an unmodified elongation useful for comparison. No. 39 was taken for anatomical comparisons.

No. 40 was 2.3 m. in height on Jan. 28, with a diameter of about 10 mm. at base, being very slender. Its age was 11 years. 14,000 old leaves were removed on June 4, after which time no elongation occurred.

The normal tree made an elongation of 131 mm., No. 34 added 56 mm. to its length after defoliation at the end of January, with cessation of growth early in June. No. 35 made an elongation of 72 mm. after defoliation at the end of February. Its length was 26 mm. on Jan. 28. No. 36 made an elongation of 15 mm. after defoliation at the end of March. No. 37 had made an elongation of 138 mm. for the season when the leaves were removed May 30. The leader of No. 38 elongated 150 mm. by the end of May, while that of No. 40 added 205 mm.

No definite notation of the development of adventitious buds was kept, but these were seen in many of the operations much as is given in the citation of results by Lutz below.

No diametral measurements were made but an examination of the larger branches in Sept. 1932 did not reveal any anatomical differences of note between the defoliated trees and No. 39, the normal specimen selected for this purpose. All trees of the series were of normal appearance and stature in June, 1938.

The effect of complete defoliation was seen to depend upon the seasonal stage at which the operation was performed. When carried out during the dormant period preceding incipient growth, death or failure of the cambium resulted in oaks and pines. The effects of removal of old leaves after the buds had unfolded were most serious in the earlier stages of development of new leaves with a lessening effect with the advance of the season.

Lutz defoliated a series of small trees of *Pinus silvestris* beginning in March before the inception of growth with the final operation late in August near the end of the season of diametral enlargement. Removal of leaves during the dormant period was followed by death. Defoliation after the new shoots had reached a length of 10 cm. was accompanied by some wood formation. Later operations produced lesser deviations from the normal. Some adventitious buds awakened after all defoliations, the maximum number coming with the earlier operations. The tracheids developed after defoliation were of maximum radial diameter but with walls less thickened than the normal.[1]

Partial Defoliation of Pines

The Monterey pine carries three suites of leaves through the growing season and discards the oldest in September and October, with the result that a much reduced foliar surface is carried through the dormant season. Removal of these leaves at this period results in the early death of the tree.

In an additional series of experiments, the leaves of the youngest suite which had matured in the previous August were removed in November. This was done in simulation of the operations of the tussock moth (*Coloradia* sp.?) which feeds on the needles of the terminal segment only leaving the buds intact (Fig. 1). Such a food selection would imply special features of chemical content which might be of importance in the promotion of growth.

[1] Lutz, K. G. "Beiträge zur Physiologie der Holzgewächse." *Beiträg. zur wiss. Bot.*, 1, 1895, 1–80.

FIG. 1. Branches of a young Monterey pine, the fascicles of three of which have been cut by the tussock moth leaving the minute buds intact. Some fascicular buds have awakened. The most profuse proliferation has taken place in the basal parts of the denuded segments and in the intact fascicles of contiguous upper region of the segment below. (Figs. 2, 3, and 4.)

In the experiments in which trees were completely defoliated the leaves were pulled from their sheaths in a manner which would inflict minimum damage on the twigs. The minute central dormant buds of the fascicles, however, were destroyed or removed in the operation. In the partial defoliation this damage was avoided by cutting the needles a few mm. above their common base (Figs. 2, 3 and 4).

In January 1939 it was seen that most of the stubs of the needles were dead, although some green bases were present. The terminal segments of the twigs which had been denuded included some which were shrivelled and dead. The terminal buds of many were alive but some were shrunken. Some buds awakened in the stubs of the excised

fascicles (Fig. 4) and many were activated in the intact fascicles of the segment below, the number being greatest in the upper part of the segment. The elongating axes bore single primary leaves.

Cambial activity in neighboring pines began late in February, but no wood formation had taken place in the small trunk of the experi-

FIG. 2. Effects of excision of fascicles on uppermost branches of young Monterey pine. The needles were cleanly cut 2–3 mm. above the base leaving the central minute fascicular buds intact, November 1938. The apical winter buds were still dormant when photographed in February 1939. Short primary leaves have been developed by the awakening buds in the bases of the excised fascicles. Similar proliferation in intact fascicles occurred in the upper part of the segment below. (Figs. 1, 3, and 4.)

mental tree by April, an effect which can be reasonably ascribed to the removal of the youngest suite of leaves.

It is to be made clear that the positive effect is that of the awakening of minute buds terminal to the arrested branch fused with the common bases of the three leaves of the fascicles. Similar proliferations in fascicles of this and other pines resulting from injuries and also appear-

ing under unusually favorable conditions for growth have been described by Lloyd.[1]

The effects in question may not be attributed solely to traumatism, and as the terminal buds were left intact no weight can be given to

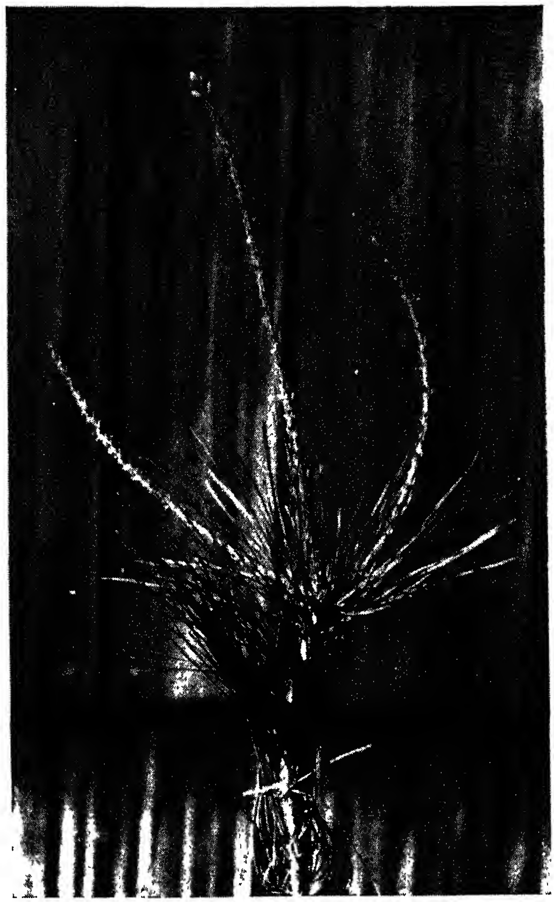


FIG. 3. Effects of excision of fascicles of all branches of a Monterey pine 20 years old. Leaves cut as in Fig. 2, November 1938, photographed Feb. 1939. Apical buds dormant. Some buds awakened in the bases of the excised fascicles. More profuse proliferation in the intact fascicles of the segment below. Cambium of the main stem remained inactive. (Figs. 1, 2, and 4.)

the supposed inhibiting action of their products. Whatever possible accelerated formation of growth promoting substances may take place it is to be noted that the primary generative cells of buds are awakened

[1] Lloyd, F. E. "Morphological Instability, especially in *Pinus radiata*." *Botan. Gazette*, 57, 1914, 314–319.

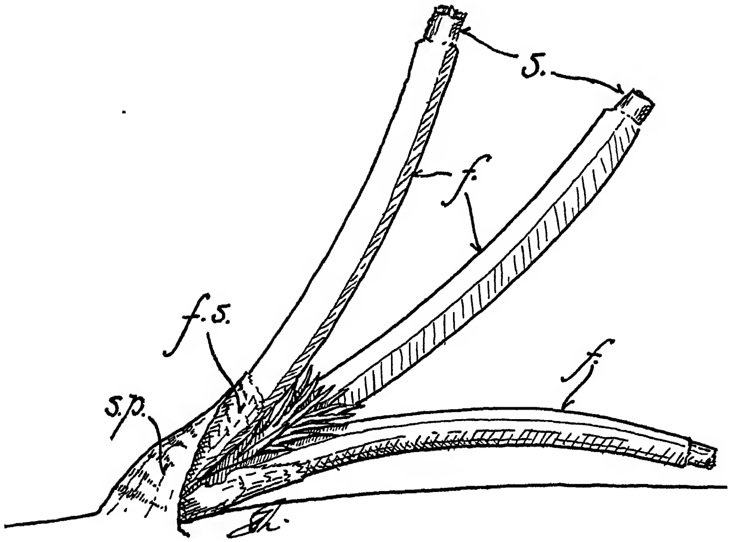


Fig. 4. Basal region of fascicle, the needles of which, *f*, *f*, and *f*, have been cut a few mm. above their bases. *sp*, bract, *fs*, fascicular sheath. The simple leaves of the central proliferating bud are to be seen. Drawn by Prof. F. E. Lloyd.

while the secondary meristem remains inactive, results which are in contrast with those of disbudding described in the following section.

The Effects of Disbudding of Pines

Jost (1891, 1893) found that removal of the terminal buds from some branches of *Pinus laricio* early in March, about the time of beginning growth, was followed by awakening of the terminal buds on shorter branches and by some cambial action, much less than the normal. Dissection of branches in the autumn from which the buds had been removed early in May revealed a layer of thin tracheids which must have been formed about the time of the decapitation. Defoliated branches retaining buds displayed cambial action, but removal of the latest suite of leaves and of the buds was followed by but little growth.[1]

A small tree, Monterey pine No. 45, on which measurements of radial growth of stem and roots had been made for several years, was selected for disbudding operations. This tree stood on a southward facing dry slope and was about 12 or 13 years old. A dendrometer was attached to the small trunk at a height of a meter from the base. All branches below this level had been discarded by the tree or removed

[1] Jost, L. "Ueber Dickenwachstum und Jahresringbilding." *Bot. Ztg.*, **49**, 1891, 485–495, etc. Jost, L. "Ueber Beziehungen zwischen Blattentwicklung und der Gefässbildung in der Pflanze." *Bot. Ztg.*, **58**, 1893, 89–138.

in preceding seasons. A dendrometer was put in bearing on the largest root at a place about 30 cm. from the base of the trunk. Scaffolding was arranged to facilitate examination of all terminals.

After the cambium had passed into a resting condition in the autumn of 1937, and the suite of leaves of 1935 had been discarded in the usual manner in September and October, all buds were broken off in mid-November 1937.

Buds in the uppermost fascicles awoke in great number in January at the time when the terminals of other trees of this species were breaking open. These newly awakened buds fell off under a very slight lateral pressure. Such removal was followed by the activation of others in the terminal region of the segment, and in some cases of a few in the segment formed in 1936. So profuse was this action that it was necessary to make a detailed examination of all branches at intervals of 10–15 days, and remove the minute buds which had been developed. No account of the number was kept during the entire period, but 400 were removed in September, at which time a few closed winter buds were noted and allowed to remain. The autumnal abscission of the oldest leaves which usually began at this time affected only a few of the suite of 1936 of the disbudded tree, which were still intact in April 1939. This unusual retention of old leaves led to some measurements which disclosed no secondary elongations of these organs (Fig. 5).

Cambial action was displayed throughout the usual period, coming to an end early in July, with a diametral increase of 4.5 mm. As accretions of 4.2 mm. in 1935, 4.2 mm. in 1936, and of 9.3 mm. in 1937, had taken place, no certain effect of disbudding is to be assumed. The season of 1937 was unusually favorable to growth to trees on this site.

Diametral increase of the root began in December 1935 and continued until the following February with an accretion of 0.7 mm. Inception of cambial activity recurred in December 1936 and continued until May 1937, with an accretion of 1.5 mm. Increase began in November and came to an end early in April 1938, with an addition of 2 mm. to the diameter of this part of the root. Inception and cessation of cambial activity of the root were seen to be earlier than in the stem as in a normal tree. No identifiable modifications of growth in the cambium of either root or shoot to be connected with disbudding were seen.

Briefly summarized the removal of all dormant terminal buds of the Monterey pine results in the awakening of the minute fascicular buds in much the same manner as after removal of the youngest suite of leaves described in the previous section. Unlike consequences in-

clude a normal activity of the cambium and the retention of leaf-fascicles for a season after the time of their normal abscission. Growth promoting substances necessary for the activity of both primary and secondary meristem may be assumed to be produced by the leaves, probably by those of the youngest suite.

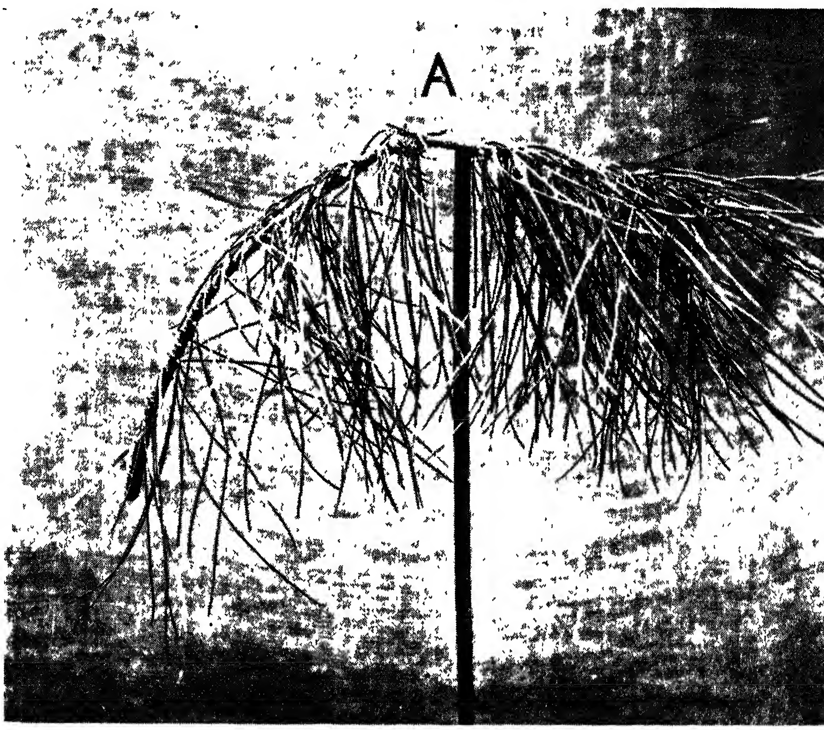


FIG. 5. Segments of a branch of Monterey pine formed in 1936 and 1937, disbudded in Nov. 1937 and photographed Feb. 1939. No elongation occurred in 1938 and the fascicles to the left of *A* developed in 1936 were not cast off in 1938 in accordance with usual procedure. Proliferation of the minute fascicular buds occurred as Figs. 1, 2, 3, and 4. Many hundreds of such awakening buds were removed, but one was allowed to form a winter bud *A*.

Defoliation of Evergreen Oaks

Experiments with evergreen oaks included defoliations of *Quercus hypoleuca* and *Q. agrifolia*. Two trees of *Q. hypoleuca* native to the mountains of Arizona, grown at sea-level, Carmel, had been kept under observation for twenty years and the seasonal habit of growth established by dendrographic measurements.

The leathery leaves emerging from the buds in April are held until the corresponding time of the following year when a new suite is developed. Elongation is at an end in about 40 days, and the youngest

leaves reach full size a few days later. Activity of the cambium which begins about the time of awakening of the buds or a few days later persists for two months after the cessation of the first growth in length. After a fortnight of quiescence of the terminals many buds open and a mid-summer extension of the twigs occurs which may not come to an end until August or September with but little evident connection with radial growth, except in the twigs near the active buds (Wight 1929).[1]

An operation to test the effect of the removal of old leaves during the resting period on the subsequent action of the buds and cambium was carried out in 1938. A total of 6,200 leaves with a plane area of 6.5 sq. m. was removed from one of the trees Jan. 12–20, 1938, about two months preceding the inception of seasonal growth of buds and of cambium. Detachment of the leaves was carried out by snipping the petioles above the enlarged bases in which the abscission layers would have been formed. The roots had shown some white tips since the preceding autumn. That no pronounced traumatic effect had been set up was evidenced by the fact that the short stumps of the petioles were cast off in the spring much after the usual manner of detachment of the leaves.

The buds of the defoliated tree showed an enlargement beginning about a fortnight earlier than the normal. The buds awakened were those which would have opened under normal conditions, and the number was not greater than in the normal tree. The new leaves developed to the usual size, reaching maturity a fortnight earlier than those of the normal tree. A similar precession of development characterized the catkins and pistils. The acorns of the normal tree soon reached a size greater than that of the defoliated one, but all except a few ceased development and not more than a dozen emerged from the cups in July, and these dropped off before reaching maturity in August. On the other hand the acorns of the defoliated tree developed steadily, the exposed part being as long as the cup by August 1st. Maturity was attained before the end of the month and some were collected, while others fell off in a normal manner.

Dendrometric measurements of the basal parts of trunks of the experimental and control trees in 1936 and 1937 included slight swellings in January and February with accretions of 4 or 5 mm. to the diameter during the summer growing season. The control tree showed an increase of 5.4 mm. during 1938. The defoliated tree showed no increase in diameter until about two months later than the normal, a

[1] W. Wight. "Secondary Elongation Growth in Oaks." *The Naturalist*, February 1930.

month after its leaves had reached full size, and the total accretion was but 1 mm. for the season. A midsummer or secondary awakening of buds occurred much as in the normal tree.

Defoliation of this oak during the resting season did not induce development of adventitious buds, nor were lateral or accessory buds awakened in any manner deviating from the normal.

The removal of mature leaves of the California live-oak, *Quercus agrifolia,* during the winter period is followed by somewhat more complicated effects. The dormancy of buds in this tree is much less positive than in *Q. hypoleuca,* as a few active buds may be seen in January, although an inclusive awakening does not take place until two months later.

Removal of leaves in the November–January period is followed by the awakening of terminal and lateral buds to a number of five or six times as great as of those removed. The immediate development of a new suite of leaves in *Q. agrifolia* equips the tree with adequate foliar surfaces by the time of normal inception of radial growth in March or April, and a layer of wood of average thickness is formed, but the component elements did not carry out secondary thickening of the walls.

Repeated defoliation in May and in July resulted in an indefinite layer of cambium derivatives in which differentiation had not proceeded very far.

Removal of the old leaves was seen to hasten the opening of buds, markedly so in *Q. agrifolia,* and only slightly in *Q. hypoleuca.* The immediate replacement of the leaves was followed by activity of the cambium of *Q. agrifolia* which resulted in a normal number of cell divisions with a failure of differentiation. The replacement of the leaves of *Q. hypoleuca* was followed by reduced cell-division, the elements reaching normal maturity.

The activity of the old leaves is seen to be necessary for the formation of a layer of wood of normal thickness in both trees. Transpiration would be reduced to a minimum for nearly three months in *Q. hypoleuca,* and for not more than a month in *Q. agrifolia.*

When the effects of disbudding and defoliation of broad-leaved deciduous trees are compared with the foregoing certain similarities become apparent. Kuhns (1910) [1] removed the nearly mature leaves of *Æsculus, Betula, Corylus, Populus* and *Sorbus* at a time when about half the average amount of wood had been laid down and saw the awakening of what was probably about an average number of buds

[1] R. Kuhns. "Die Verdoppelung des Jahresringes durch Entlaubung." *Bibliotheca Bot.*, **15**, Hft. 70, 1910.

and the formation of a second suite of leaves. Xylem formed after defoliation comprised elements with thin walls and imperfectly defined parenchymatous elements.

Disbudding of a beech tree, *Fagus silvatica*, by Lutz [1] on March 20, before the inception of growth, was followed by no growth in diameter, but with a continuous emergence of dormant or accessory buds until the end of the season, and with the development of winter buds which were normal.

Removal of newly formed leaves in May and of awakening buds throughout the season was followed by the addition of some wood but the tree died at the end of the season. A similar operation in mid-June resulted in stoppage of diametral growth, awakening of dormant buds and death at the end of the season. Later defoliations caused less deviation from the normal, and were not followed by death. It appeared that the cancellation of foliar activities was as in the oak and pine inimical to growth. It is to be noted that diametral growth in Fagus does not begin until leaves have attained nearly full size (MacDougal, 1938).[2]

In intensive observations on *Acer*, *Æsculus*, *Fraxinus*, *Malus*, *Populus*, *Platanus* and other trees it has been found that the growth promoting substances are to be found in greatest concentration during the growing season in twigs and small branches, with lesser maxima in the trunk immediately below the crown, in the unified bases of the trunk and large roots and in root tips.

The amount of such substances is greatest shortly after the inception of seasonal activity after which a decrease occurs in consequence of which only negative results are shown by tests during the winter dormant or resting period. An abrupt appearance and rise in concentration marks the beginning of the growing season. A graph representing these variations is similar to one expressing the amount of starch present during the same period. The disappearance, conversion or hydrolysis of starch and its reappearance, coincidentally with growth promoting substances, is a concurrence which adequately analyzed may yield much information of value in the interpretation of the seasonal cycle, especially in evergreen trees.

From the results of the observations described in the preceding paragraphs it is clearly apparent that the origination of the auxins

[1] K. G. Lutz. "Beitrage z. Physiologie d. Holzgewächse." *Beitr. z. wiss. Bot.*, 1897, 1–80.

[2] D. T. MacDougal. *Tree Growth*. Leiden, 1938, p. 160. Avery, Burkholder and Creighton. "Production and Distribution of Growth Hormone in Shoots of *Æsculus* and *Malus*, and its Probable Role in Stimulating Cambial Activity." *Amer. Jour. of Bot.*, 24, January, 1937, 51. H. Söding. "Wuchstoff und Kambiumtätigkeit der Baüme." *Jahrb. f. wiss. Bot.*, 24, Hft. 4, 1937, 639–670.

necessary for cambial activity in trees is to be attributed chiefly to the leaves. It is also obvious that correlative effects, such as inhibition of lateral buds, must be assigned mainly to free auxins translocated from terminal buds within comparatively short distance. Such effects do not come exclusively from buds, however, as was evidenced by the awakening of dormant buds following partial defoliation.[1]

That basipetal diffusion or translocation of growth promoting substances in twigs and branches identifiable by microscopical evidence with extension of cambium activity takes place is well established. That a gradual extension down the main shaft of the tree follows is not supported by observations. In one tree, *Parkinsonia,* many seasons may pass with no cambium activity of the trunk although elongation, thickening, development of leaves and flowers may take place. In other trees inception of cambial activity may be found at various levels from the base to the lower level of the crown, or one flank only.

In the case of deciduous trees which show cambial activity previously to the opening of the buds with no foliar surfaces exposed since the previous season, the possibility that cambium derivatives incompletely differentiated, may resume enlargement by the action of residual auxins under favorable temperatures is to be taken into account.

In evergreen trees, such as the conifers and oaks, a continuous production of growth promoting substances might be accompanied by their continuous translocation and accumulation throughout the body of the tree as "bound auxin" so that growth would follow the advent of favorable temperatures (Went).[2] The prevalence of such conditions would make possible the simultaneous awakening of the cambium throughout the entire length of the trunk as has been observed many times. No translocation of growth promoting substances downward at a rate exceeding 10 mm. per hour has been found. When these substances enter the transpiration stream as in the roots they might be carried from the place of origin to the summit of the crown within a day (Both 1937, and Dixon 1938).[3] In any case the conception of cambial action as being the direct effect of downward translocation of auxins or of their stimulative action is no longer tenable.

[1] W. A. Zimmerman. "Untersuchungen uber die raumliche und zeitliche Verteilung des Wuchsstoffes bei Pflanzen." *Zeitschrift f. Botanik.*, 30, 1936, 211–252.

[2] F. W. Went. *Remarks about Two Auxin Problems*, 4, No. 6, 1938, 503.

[3] M. P. Both. *Stoffwanderung in einfachen Systemen.* Inaug. Diss. Groningen, 1937. H. H. Dixon. "Transport of Substances in Plants." *Proc. Roy. Soc. Lond.*, Ser. B, No. 838, 125, March 1938, 1–25.

EFFECTS OF GIRDLING

Girdling by the removal of a belt of bark, phloem and cambium from a woody stem, used in practice to kill trees in clearing lands for agricultural purposes, severs the conduits by which leaf-products are transmitted to the basal regions and to roots. This operation may not seriously disturb the upwardly moving mesh of sap immediately although it may do so in a few days, or a few weeks, a matter dependent upon the structure and arrangement of the conduits. The greater part of the ascent of sap is in a few outer layers of the xylem, but some movement may take place in the hydrostatic meshwork in old layers. Reduction of the conduction capacity of the outer layers may be followed by increased movement in the inner layers, even in the heartwood as in *Sequoia* described below.

Girdling of trunks of *Quercus agrifolia*, the California live-oak, is followed by the formation of a callus above the denuded zone, from which callus bridges may be extended downward. No growing points or buds are differentiated, but a cambium is finally formed. Dormant buds below awaken and long sprouts are formed with increase of root-sprouts. Sap is conducted upward in many layers and radial movement through the ray-tracheids make possible extended survival after girdling. Callus is rarely formed on the lower margins of girdles, but is developed on stumps of branches and trunks, and growing points are differentiated, which may give rise to sprouts.

Redwood trees (*Sequoia*), in their prolonged endurance after girdling, furnish some important information as to the possible upward conduction of leaf-products including growth promoting substances. The dendrographic record of a tree which survived 8 years after girdling of the trunk has been previously published.[1] More extensive observations were made on a group of large trees which were girdled in 1920 and kept under observation until 1939. One tree, the root-system of which was injured by road making, died within three years. Two others formed callus bridges across the greater part of the girdles and developed strong root sprouts and resumed growth in a normal manner. A fourth (27A) was felled for examination in 1937 and two others were alive in 1939. Both displayed growth activity in the 20th season after girdling. All had developed root-sprouts, which in two cases had attained the stature of small trees.

A band of bark, phloem and cambium 15–20 cm. in width had been cleanly removed in girdling in 1920. The exposed sapwood when ex-

[1] D. T. MacDougal. "Studies in Tree Growth by the Dendrographic Method." *Publ. 462, Carnegie Inst. of Wash.*, 1936, p. 230.

amined in 1938 was seen to consist of 20–30 layers of dried and brittle cells. The reddish heartwood was moist and evidently carried the ascending sap. The details of one tree (No. 27C) may be cited as illustrative of what had taken place. The diameter of the trunk including bark above the girdle was 97 cm., below 90 cm., while the woody cylinder of the girdle was 90 cm. by the diameter tape (Fig. 5).

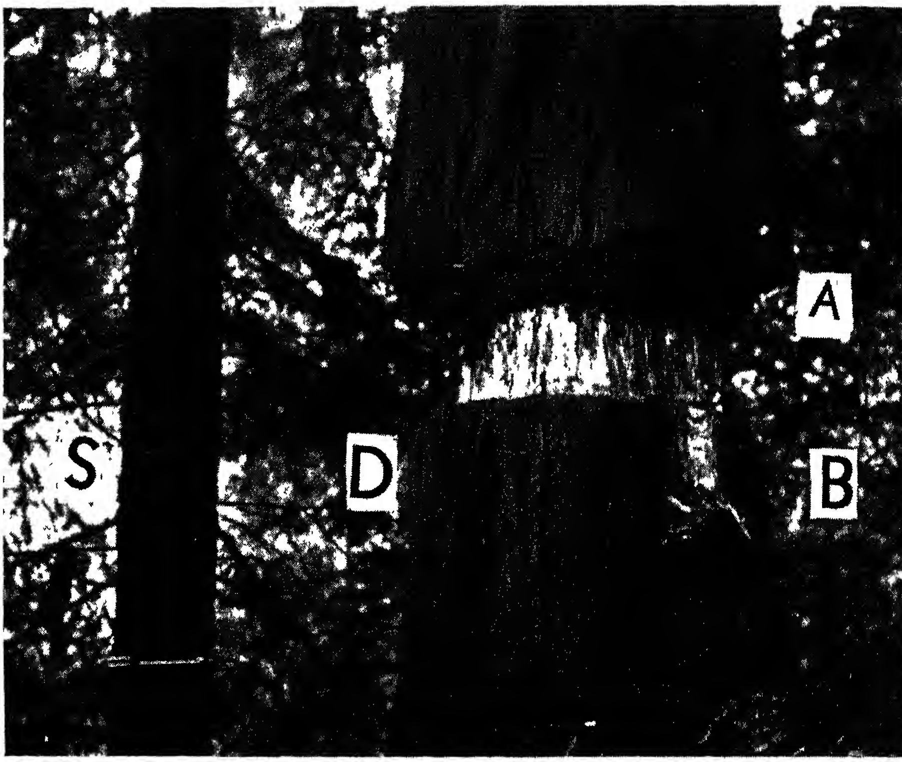


FIG. 6. Girdled trunk of large *Sequoia* with small trunk at left which arose as a root-sprout after girdling, in 1920. Photographed Feb. 1939. A heavy callus has been formed above the girdle (*A*), but none below. Bark on right flank at *B* exfoliating. Active cambium on opposite flank at *D* maintained by leaf-products from small trunk. Conduction of nutritive material and of growth promoting substances downward in small trunk and upward in root and main trunk is implied. Conduction through girdle could take place only in heartwood as the 20–25 layers of sap-wood are dry and brittle.

The bark was separating from the wood below the girdle except in a sector comprising about one-fourth of the circumference which faced the only root-sprout present. This now had a basal diameter of 40 cm. About 20 layers were discernible in the small trunk. The wood formed on the large trunk below the girdle, and after the operation, included 19 layers which were two or three times the thickness of the layers

above the girdle. Two layers may have been formed in some years, and it is reasonable to assume that cambial action was resumed in this sector of the main trunk within a few years after girdling.

The entire arrangement implies that leaf-products of the root-sprout were conducted from its base a short distance in a root to the trunk and thence upward to a height of 2 meters to be used in the construction of new cells by cambial activity. While translocation upward in the phloem would be as rapid as it might be downward, it is also to be said that any organic material diffusing into the sap stream would be carried upward at a much higher rate and would be subject to radial withdrawal at all levels by the osmotic action of the phloem and cambium (Dixon).[1] Upward conduction of growth promoting substances is implied in both cases, as it is also in the results obtained by Söding from the girdling of a forked tree of *Tilia cordata.*

Other possibilities of the upward conduction of organic material in leafless sprouts have been cited by Weevers. The chief interest of such movement in connection with the present theme lies in an implied upward translocation of growth promoting substances, which in general have been found to move downward only.

Girdling operations with the Monterey pine also yield some evidence that the cambium below the denuded zone may not die the first season but may be thrown into an inactive condition from which it may recover if a callus bridge across the girdle is formed. It was also found that if the girdle was made above leafy branches, while no callus was developed on the lower margin, yet the leaf-products participated in the prolonged endurance and survival of the tree. In no case was the upward conduction so marked or over such comparatively great distances as in *Sequoia.*

It can safely be assumed that stoppages of growth below a girdled zone cannot be attributed to an exhaustion of such main building materials as fat and starch, although the relative amounts of these substances may be modified as a result of the operation. Thus Söding [2] found that all times relatively more starch was to be found in the xylem and phloem of *Tilia* above than in these regions below the girdle. At the same time as much or more fatty material was found in the phloem below the girdle than above.

[1] H. H. Dixon. "Transport of Substances in Plants." *Proc. Roy. Soc. London,* Ser. B, No. 838, **125**, March 1938, 1–25.

[2] H. Söding. "Physiologische und anatomische Untersuchungen an einer geringelten Linde." *Planta,* **4**, Hft. 3, 1927, 373–388. O. F. Curtis. "Studies on the Tissues Concerned in the Transfer of Solutes in Plants." *Ann. of Bot.,* **39**, 1925, 573–585. Th. Weevers. "Die Ergebnisse einiger Ringelungsversuche und ihr Bedeutung fur die Stoffwanderung." *Recuil. Travaux Bot. Neerland.,* **25** A, 1928, 461–474.

Starch above the girdle remained less completely liquefied in a normal winter procedure than in a normal trunk. Hydrolysis of starch occurred earlier than in the normal below the girdle, and restoration of this material in the spring was belated. Development of conducting elements in xylem and phloem were seen to be impeded both above and below the girdle, while parenchymatous elements were exaggerated. Similar conditions have been recorded by Russow, Mer. Fischer, Muller, Krieg, Hibino, Th. Hartig, LeClerc du Sablon and others in *Quercus*, *Betula*, *Cornus*, *Populus*, and many other species.

Discussion

Removal of a zone of phloem and cambium in all cases breaks the conduits by which the greater part of leaf-products from the crown would reach the region of the trunk below. A progressive deterioration of the outermost conduits in the xylem by which solutions pass upward in the trunk would also ensue. The seriousness of the effects would vary with the anatomy of the trunk from species to species. If the greater part of the sap passes upward through the outermost layers of wood and if radial conduction is slow by reason of the structure of the wood then girdling will be followed by death within a season or two from lack of adequate water supply to the crown.

If on the other hand sap may pass upward freely in old wood, and if the structure is suitable for the radial movement of sap, as in the rays of *Quercus*, *Salix*, *Sequoia* and others, then the water supply may be maintained in such manner as to prolong the existence of the tree as exemplified by *Sequoia*.

While nearly all identifiable movements of growth promoting substances or their effects have been found to be in a basipetal direction, girdled redwoods offer indubitable illustrations of upward conduction of leaf-products derived from root-sprouts. Organic material adequate for the formation of comparatively heavy layers of wood may be carried upward in the xylem or phloem. This continued translocation takes place in trunks in which the entire supply of sap of large trees is conducted in the heartwood past the girdles, after which it moves radially through 40 or more layers of wood outward to the cambium.

Notable progress has been made in the determination of the general nature of substances which will activate generative cells. The manner in which their effects are extended or propagated is as yet undetermined. Positive conclusions may be reached only after localization of origin of these substances and identification of accessory substances.

Upward conduction for a distance of 2 meters is implied in the reactions of girdled *Sequoias*. Basipetal extension of incipient growth has been followed only to the bases of twigs and branches and of small shoots by any observer. That other conditions prevail lower and in the main trunk is exemplified by *Parkinsonia*, in which elongation and thickening of the twigs occurs in every season, but several seasons may pass with no discernible cambial action of the trunk.

Similar restriction of growth to twigs and branches and to only a short distance from the terminals is seen in the secondary, or midsummer growth of oaks.

In one series of measurements registering dendrometers and dendrographs were attached at two levels on the trunk of a Monterey pine and at four places on a root 30 meters in length. The records of several years include no instances of progressive extension of growth in trunks, from trunk to root, or in roots in either direction. It has been found that the activity of the primary generative cones in the tips and of the cambium may begin or end independently and with varied precedence. These facts as well as the identification of growth promoting substances in root-tips which respond to the same tests as those from buds and leaves modify materially any hypothesis as to extension of effects from shoots into roots.

The region comprising the basal flares of trunks and the thickened bases of attached roots suggests the conjoint action or additive effect of growth promoting substances from both trunks and root-tips. This conception is not to be confused with the suggestion of Cockerham [1] that the "downward spread of xylem activity is superposed on the normal root activity in the middle region of the roots" (of *Acer*). High concentration of growth promoting substances have been found in the united bases of trunks and roots. The possible influence of the almost constant flexion to which the growing cells are subject due to wind action might account for some of this amount and for increased cambial action.

Wood-formation is greater on the upper stretched side of a dicotyledonous stem laid in a horizontal position than on the lower side which is compressed. The reverse is true of coniferous stems. Exceptions are seen in both categories. These differential effects may reasonably be ascribed to differences in cambial procedure as has been proposed by Priestley and Tong.[2] In addition to the part played by gravity the

[1] G. Cockerham. "Some Observations on Cambial Activity and Seasonal Starch Content in Sycamore (*Acer pseudo planataus*)." *Proc. Leeds Phil. Soc.*, 2, Pt. 2, 1930, 64–80.

[2] J. H. Priestley and D. Tong. "The Effect of Gravity upon the Cambial Activity in Trees." *Proc. Leeds Philosoph. Soc.*, 1, Pt. 5, 1927, 199–208.

possible effects of continuous stretching on one flank and of compression on the other are to be taken into account.

Eccentricity is reported not to appear in roots fully submerged in the soil but is manifested in portions laid bare, which is suggestive of the influence of light on the growth promoting substances. It has long been known that in roots bent by mechanical means or by geotropic response laterals originate in the stretched pericycle of the convex flank. Whether the formation of the new root-initials may be attributed to a different form or state of auxin from that which activates the cambium cannot be stated. The physical conditions of translocation of growth promoting substances originating in root tips are notably different from those prevalent in stems. Downward movement of leaf-products in the phloem may not exceed a rate of a few mm. per hour in a direction opposite to that of the transpiration stream. Substances in solutions in the roots may be carried the entire length of the longest roots or even to the summit of the crown within a day.

Upward movement of leaf-products in leafless sprouts and inflorescences has long been known. Translocation takes place in some instances at a rate and to an amount greater than possible by osmotic action. It is also a widely accepted fact that at the time of awakening buds and inception of cambial action that sugars presumably derived from starch which disappears during the winter resting season in many trees have diffused into the xylem to a notable concentration and would be inevitably carried up by the sap stream in the xylem. Other organic substances would be moved at the same time. Radial movement would follow the osmotic gradients.

Such a pattern might be followed by leaf-products including auxins from root-sprouts below a girdle in *Sequoia.* At this level radial movement would be from the outer layers of sapwood directly to the cambium and phloem. As the sap-stream passes the girdle only in heartwood radial conduction above would pass 20 to 40 seasonal layers of wood to reach the cambium.

Recapitulation

The principal results of the experimentation described in the preceding pages may be briefly summarized as follows:

1. Disbudding of Monterey pine in the dormant season (November) is followed by awakening of growing points in the axils of the leaf-fascicles of the segments formed in the two previous seasons. Removal of the minute buds as formed is followed by the appearance of others during the entire year.

2. Removal of buds as above does not affect the activity of the cambium in branches, trunk or roots in two subsequent seasons.

3. The oldest suite of leaves which would normally be cast off in the following autumn (ten months later) was retained.

4. The facts enumerated in 1, 2, and 3 support the assumption that removal of the buds cancelled the inhibitory action of bud-products and the abscission processes at the bases of the leaf-fascicles but did not destroy any substances necessary for cambial activity. Otherwise expressed, substances originating in the buds affected primary generative cells only.

5. Defoliation of the Monterey pine in the dormant period resulted in the death of the tree in the following season, with no previous awakening of either primary or secondary generative cells.

6. Complete defoliation of the pine after awakening of buds and beginning development of leaves resulted in some modification of the amount and structure of the wood formed, the effect diminishing with deferment of the operation.

7. Partial defoliation, by removal of the youngest suite of leaves, including about three-fifths of the total number, in the dormant season results in the death of many of the denuded segments and of their terminal buds. Buds may develop in the axils of the excised fascicles, and in the axils of the fascicles of the segment below. The surviving terminal buds display a stunted development of leaves in the following season. Activity of the cambium begins in the season following. These results make it evident that active auxins which may inhibit development of the buds are produced in young leaves as well as buds; also that products of old leaves during the dormant period may activate cambium.

8. Removal of all leaves from two species of evergreen oaks in the dormant period did not kill the trees as in the pine, but the activity of the cambium in the following season was greatly restricted and the walls of the elements of the xylem did not complete secondary thickening.

9. The undamaged buds of the oaks opened slightly in advance of the average date and full leaf-surface as well as adequate surplus food-material were present many weeks before the beginning of the growing season.

10. It may be reasonably assumed that some material necessary for cambial action is produced in the leaves of oaks and, or, that the transpiratory action of the evergreen leaves brings up material in the xylem necessary for growth.

11. Girdling operations on old trunks of Monterey pine are followed by death of the tree in the next season unless callus bridges across the girdle are developed.

12. Girdling of small pine trees above branches may affect growth only temporarily.

13. Gradual extension of cambial activity from near buds downward in twigs and branches has been repeatedly observed in many trees but a similar continuation down the trunk in any tree has not been established.

14. Apical and cambium activity of twigs of oak in midsummer, or secondary growth, does not extend to trunks. A desert tree, *Parkinsonia*, may display elongation and thickening of twigs for many years in succession with no cambium action of green-barked trunk.

15. The flaring bases of trunks and the basal sections of attached roots display unified action and have the appearance of being the seat of additive effects of growth promoting substances from both roots and trunk.

16. The basal region of a tree and of attached roots is subject to constant flexion. Alternating stretching and compression might be expected to have effects similar to those found in horizontal branches and exposed roots.

17. Long continued observations have uncovered no evidence that the effects of the growth promoting substances in trunks extend beyond the neighboring basal regions of attached roots.

18. Apical growth of roots may precede by weeks or even months. Activity of the shoot and cambial activity of these organs which follows may not be simultaneous with cambial action of trunks.

19. Formation of initials from which lateral branches of roots may be developed involves the pericycle and cambium and appears to be promoted by substances originating in roots. Such initials may also arise by the action of geotropism, or by mechanical curvature, etc.

20. Development of branches by the shoot consists in the activation of primary generative cell-masses, on which terminal buds are seen to exert an inhibitory effect.

21. Survival of a tree after girdling depends upon the availability of leaf-products below the girdle, capacity for conduction in the modified wood of the girdle, and radially above the girdle.

22. *Sequoia sempervirens* furnishes an example of prolonged survival after girdling. Three trees girdled in 1920 lived 18–20 years. During the greater part of this period upward conduction past girdles was

possible only in heartwood 30–40 years old and radial conduction to the cambium through 40 or more layers of wood.

23. The term leaf-products is used in the previous paragraphs to include nutritive material and growth promoting substances originating in chlorophyllose tracts. While their general translocation downward takes place in the phloem, they and other organic substances diffusing into the xylem may be carried upward in the sap stream in notable quantity at the time of awakening buds and unfolding leaves. Leaf-products of root-sprouts may reach cambium below a girdle in this manner.

DEVELOPMENTAL EFFECTS RESULTING FROM EXPOSURE TO X-RAYS

I. Effect on the Embryo of Irradiation of Frog Sperm

ROBERTS RUGH *

Columbia University and New York University †

(Communicated by Edwin G. Conklin)

ABSTRACT

Uniform suspensions of sperm from hibernating frogs, *Rana pipiens*, were irradiated with x-rays at 200 kv. from 15 *r* to 50,000 *r* prior to using them for artificially inseminating frog's eggs. No effect on motility or fertilizing powers of sperm was detected; nor on rate or pattern of early cleavages. Even at 15 *r* some resulting embryos showed exo-gastrulation and spina bifida while at 1,000 *r* some perfectly normal embryos resulted. This discrepancy is explained as a result of chance bombardment of vital points in chromosomes, since at higher doses all resulting embryos are abnormal. If hatching is considered as a critical end-point, there is a typical exponential curve of hatching percentage dropping from 97.8 per cent at 15 *r* to 1.6 per cent at 10,000 *r* and rising again to 90.5 per cent at 50,000 *r*. This latter rise is explained as due to incapacitating of sperm for syngamy but not for activation of the egg; hence gynogenetic development follows. Such abnormalities as are produced have been produced by other means such as subjecting the developing embryo to drastic environmental changes. Here the effect is applied through the sperm by irradiation prior to insemination of normal eggs. Preliminary tests indicate disruption of chromosomes, and therefore the effect is described as due to induced lethal mutations. These experiments are foundation for further research along the lines of wave-length effect; chromosome effect; tissue culture and transplantation studies.

INTRODUCTION

SINCE the discovery of x-rays by Roentgen (1895) and of radium by Madame Curie (1898) scientists have turned to these very penetrating radiations for a solution to the enigma of cancerous growth. More recently these radiations have been intensively studied for their effects on normal biological systems. The pioneer workers in the field were severely handicapped by the lack of precision apparatus for the accurate determination of dose and range of wave-length used, and those who were particularly interested in the effects of x-rays on embryos were further handicapped by their dependence upon the natural breeding seasons. The rapid perfection of x-ray apparatus and calibrating equipment, and the recently developed technique of inducing ovulation by anterior pituitary hormone stimulation, have eliminated these two barriers. It is the purpose of this paper to lay the foundation for an intensive study of the effect on embryos resulting from the exposure of the gametes to x-rays prior to syngamy.

* Aided by a Grant from the Penrose Fund of the American Philosophical Society.

† The research of this paper was completed at Columbia University.

Since 1903, when Bohn floated a glass tube of radium bromide on water above amphibian eggs and produced gross abnormalities, there have appeared infrequent and divergent reports on the effect of radiations on amphibian material. While the major contributions have been made by the Hertwigs and more recently by Woskressensky, a summary of most of the amphibian work in this field is given below in tabular form.

a. Exposure of embryo to radiations.

1. Retardation of growth by radium (Bohn '03); and by x-rays (Baldwin '19, '20, '21).
2. Acceleration of cleavage (Hoffman '22).
3. Initial acceleration and subsequent retardation of cleavage resulting in the production of abnormalities (Gilman and Baetjer '04).
4. No acceleration of cleavage (Ancel and Vintemberger '25 and '27).
5. Production of abnormalities by radium (Bohn '03); by x-rays, indicating that abnormalities are greater when the embryo is exposed than when either gamete is irradiated (Hertwig '10, '11, '16).
6. Gastrula stage the most sensitive to irradiation (Ancel and Vintemberger '25 and '27).
7. Production of exo-gastrulae (Curtis, Cameron & Mills '36).

b. Exposure of sperm to radiations.

1. Sperm lose their motility; their capacity to fertilize; and when they do stimulate the egg, development usually ceases at gastrulation (Bardeen '07).
2. Exposure of sperm to x-rays did not markedly modify their powers of fertilization (Bardeen '09).
3. Irradiated sperm give rise to the same types of abnormalities as do irradiated eggs (Bardeen '11).
4. Fewer abnormalities are produced by the irradiation of sperm than by the irradiation of embryos (Hertwig '11).
5. Irradiated sperm produced a greater percentage of abnormalities than eggs irradiated with the same dose (McGregor '08).
6. The longer the radiation the fewer abnormalities will be produced, *i.e.*, inverse ratio (Hertwig '11).

c. Exposure of the egg to radiations

1. Eggs more resistent to radiations than sperm (McGregor '08).
2. Eggs radiated in the ovary will prevent ovulation (Bardeen '09).

3. Eggs exposed to radiations cause the same embryonic abnormalities as those produced by irradiated eggs (Bardeen '11).
4. Irradiated eggs produce fewer abnormalities than irradiated embryos (Hertwig '11).
5. Abnormalities appear in inverse ratio to radiation (Hertwig '11).

d. Explanations of the biological effect of radiations.

1. Damage to chromatin of rapidly growing tissues (Bohn '03); (Hertwig '11, '16).
2. Inhibits mitosis and regeneration (Schaper '04; Levy '06).
3. Lecithin hypothesis, suggesting that yolk is peculiarly susceptible to radiations and will break down into the toxic substances choline and trimethylamine (Schwartz '03; Schaper '04; Olivieri '29).
4. Lecithin hypothesis refuted by Hertwig ('11). Emphasis on chromatin fragmentation.
5. Effects on the cytoplasmic constituents.
    a. Enzymes (Gager '08).
    b. Achromatic figures and production of autolytic enzymes (Packard '13).
    c. Fertilizin, in egg cortex (Richards et al '14, '15, '19).
    d. Denaturation and coagulation of proteins, cytoplasmic and nuclear (Janet Clark '36).

A thorough study of the effect of radiations on the Axolotl egg and a critical analysis of amphibian radiation work are given by Woskressensky ('28).

A survey of these experiments impresses one with the special value of amphibian material in radiation work and of the need for more thorough and carefully controlled experimentation and analysis. The foundation for such work is laid in this present paper through a study of the effect on the embryo of irradiation of the spermatozoa prior to insemination.

Materials and Method

Frog spermatozoa were obtained by the dissection of the excised testis (of *Rana pipiens*) in Holtfreter's modification of Ringer's for amphibia. The spermatozoa of this species, *Rana pipiens*, are mature from early fall until the normal breeding season in March. Freshly caught hibernating frogs from the same source were used in all of this work.

The sperm suspensions were usually made from ten frogs, one testis from each frog supplying spermatozoa to be irradiated while the sperm

from the other testis was used as control material. The sperm suspensions were very concentrated, representing 10 gonads in 10 cc. of modified Ringer's. The suspensions were kept in celluloid, circular fly boxes (covered) which are used in radiation experiments. These boxes measured 2.25 cm. depth and 6.8 cm. inside diameter.

Female frogs were induced to ovulate by the method originally described by Wolf ('29) and by the procedure outlined by Rugh ('35). Following anterior pituitary hormone injection the female frogs were kept at 14°C. for 48 hours prior to the time of the radiation experiment. At this temperature the frogs' eggs may be kept in viable condition for a period of a week even though they may have descended to the uteri. It is quite possible to strip from the uteri of a sexually stimulated female approximately 50–100 eggs into finger bowls where they may be artificially inseminated. A greater number of eggs per finger bowl is too crowded to allow of normal development. In order to increase the quantitative data, however, it was the usual practice to strip such batches of eggs from several females into as many finger bowls, each containing the spermatozoa from a single radiation experiment. In this way any variation in eggs from different females could be checked.

A Coolidge type, water cooled x-ray tube was used. This was operated at 200 kv. and 30 ma. No filter was used and at the point where the test material was placed for treatment (32.5 cm. from the center of the target) the output intensity (measured in air with a small celluloid ionization chamber) was 400 *r* per minute. This apparatus was available through the courtesy of Dr. Paul Henshaw of The Memorial Hospital in New York City.

After each exposure to x-rays, the doses ranging from 15 *r* to 50,000 *r*, approximately 1 cc. of sperm suspension was removed to each of the finger bowls containing 3 cc. of the modified Ringer's and to which eggs were to be added. Within 15 minutes mature eggs were stripped from an ovulating female into the sperm suspensions and the egg masses were gently shaken to insure proper exposure of all eggs to the irradiated sperm. After another 15 minutes the egg masses were flooded with the Ringer's solution and allowed to stand while the jelly covering took on water and expanded. Within several hours the eggs were gently cut from each other as far as it was possible to do so without damage. These are the regular laboratory procedures for securing normally developing frog embryos.

In order to determine whether there might be any variation in the number or types of abnormalities if the rate of development is retarded, the eggs were usually divided into two lots, one group being kept at the

laboratory temperature of 23°C. ± 1.0° and the other kept in a constant temperature room at 14°C.

The development of all eggs was watched under binocular magnification and photographs were taken of both the normal and the abnormal types as development proceeded. These photographs were taken with a Mifilmca microphotographic camera fitted, by a brass collar, into the ocular portion of a regular binocular microscope. This gave the desired low magnification which was impossible with the regular microscopic equipment and allowed photographic records during examination. In using a regular microscope lamp with intervening round-bottom water flask for the absorption of heat it was possible to take pictures on micropan film with 2 to 4 second exposures without any damage to the developing embryo. In all cases Edwal 12 developer was used to get fine grain definition. Since all of the pictures were taken with the same binocular set-up, uniform enlargement of the pictures made it possible to determine relative size of embryos from the photographs. The photographic record represents, therefore, the range of abnormalities produced while the tabulated data refer only to the arbitrarily chosen critical end-point, *i.e.*, hatching.

Experimental Data

Effect of X-rays on Spermatozoa

There was no apparent alteration in the motility or in the fertilizing powers of spermatozoa exposed to x-radiations from 15 *r* to 50,000 *r*, as compared with the controls. At all radiation doses normal insemination and egg rotation occurred in approximately 100 per cent of the eggs. Even after 24 hours at refrigerator or laboratory temperatures the percentage of fertilization was approximately the same for the radiated and the control sperm suspensions. In both cases there is marked decrease in fertilizing powers of sperm in suspension, but the results are the same for radiated and control sperm suspensions. Such radiation effects on sperm as may exist immediately after exposure in no way affect their fertilizing powers. The motility observations were microscopic only and did not include measurement of distance swimming of spermatozoa.

Effect on Cleavage of X-radiation of Spermatozoa

It was impossible to detect any alteration in either the rate or the pattern of cleavage in eggs inseminated by irradiated spermatozoa. This was equally true of the sperm exposed to 15 *r* and those exposed to 50,000 *r* where there is some evidence that the sperm stimulates the

egg but does not fuse with the egg nucleus, *i.e.*, gynogenetic development. In any normal (control) clutch of eggs there will be variation in rate of development due, no doubt, to metabolic differences affected by such environmental variables as proximity of other embryos and the supply of oxygen. Such variations occur in nature, in any clutch of eggs. In these experiments the environmental variables were reduced by the early separation of eggs, and consequently the variations in developmental rate in any group of eggs were much reduced. The onset of the first cleavage and the pattern of the early cleavages appeared in the experimental material exactly as they did in the controls. If there are any variations in developmental rate, effected by the irradiation of sperm, they are extremely small and may be detected only by exceedingly careful cell counts with embryos kept in isolated containers at very accurately controlled temperatures. Such studies are contemplated because of reported effects on cleavage (rate and pattern) in other forms.

Effect on the Gastrula of X-radiation of Spermatozoa

Even at late blastula stage there was no outward sign of any damage to the embryo by the radiation of the spermatozoa, but coincident with the formation of the dorsal lip (the first in-turning of the cells of the late blastula) there occasionally appeared the first sign of irregularity even at the lowest experimental doses used. The dorsal lip was not smooth, certain cells being delayed considerably in their inward migration. Frequently the dorsal lip failed to develop properly as the yolk seemed to bulge forth, giving rise to exo-gastrulæ and subsequently to hemi-embryos. *Delay in the formation of the yolk plug constituted, therefore,* the first indication of some organic disruption which may be associated with the damage of the sperm cell. This delay in gastrulation (yolk plug formation) was not at all frequent in the embryos whose sperm were exposed to the lower experimental dosages, but as 1,000 *r* was approached the exo-gastrula picture occurred more and more frequently and fewer of the embryos developed beyond this stage until at 10,000 *r* practically all of the embryos died as gastrulæ. It appeared, then, that with increasing exposure of the sperm to x-radiation, the lethal end-point of development for the majority was brought to an earlier stage in development but never earlier than gastrulation (see Pls. I, II, III). The great variability, to be explained later, made it impossible to gather quantitative data at these earlier stages.

Effect on the Neurula of X-radiation of Spermatozoa

Any disruption of the normal gastrulation will affect the neurula, particularly when the disruption is sufficiently great as to cause exo-

gastrulation. Hemi-embryos appeared frequently with the anterior half of the neurula apparently quite normal. This would indicate normal invagination of the roof of the archenteron. When the abnormal yolk plug was almost completely incorporated within the embryo, spina bifida often occurred (see Pls. I, II). The effect on the neurula and the tail bud stages was most marked, *i.e.*, in the greatest number of embryos, when the sperm were exposed to 500 *r* (Pl. II). Some embryos survived what appeared to be minor abnormalities at neurulation only to die at hatching or even subsequent to hatching.

Effect on Hatching of Embryos when Spermatozoa are Exposed to X-rays

All embryos which reached the hatching stage were able to emerge from their jelly capsules. This means that the radiation of spermatozoa has no effect on the hatching enzyme produced by the developing embryo. Some embryos hatched at all radiation levels except 10,000 *r* where the number ranged from 0.7 per cent at 23°C. to 2.4 per cent at 14°C. Normal hatching time at 23°C. is approximately 4 days and at 14°C. is approximately 8–9 days. While hatching occurs quite late in embryonic development, the achievement of hatching may be regarded as indicative of a certain level of development and since exo-gastrulation, hemi-embryos and spina bifida occurred with innumerable variations, hatching was chosen as the critical end point in development for quantitative data.

A brief reference to Pls. I, II, III should be made here. A photographic record was made of all abnormalities that occurred in all experimental material. These Plates attempt to show graphically only the *range* of development from the normal to the abnormal, without any attempt to suggest percentages. In each set the extremes are shown, and since no apparent abnormalities occur prior to gastrulation there are no photographs of the early cleavage stages except in the control series heading the first Plate. A more detailed explanation accompanies the Plates.

In Tables I, II, IV below an attempt is made to show both the total number of eggs inseminated and the number of embryos that hatched. These data are represented as follows: number of embryos hatched in proportion to number of eggs inseminated, as, for instance, 59/65. The data are summarized in the percentage column.

If the hatching percentage is considered at the two temperatures, 14°C. and 23°C., it will be seen that the differences are irregular and quite within the limits of experimental error in most cases. That is, retardation of development by lowering the temperature did not alter

TABLE I

EXPERIMENTAL DATA: EMBRYOS AT 23°C.

Irradiation of Sperm Roentgen Units	No. Hatched	No. Inseminated	Total	Percentage Hatched
Control. . .	100/100	67/70	167/170	98.3%
15 *r*.	100/100	80/85	180/185	97.3
25 *r*.	59/65	92/92	151/157	96.1
50 *r*.	28/34	72/72	100/106	94.3
100 *r*.	16/22	66/76	82/98	83.6
250 *r*.	18/35	92/105	110/140	78.6
500 *r*.	11/24	40/68	51/92	55.4
1,000 *r*.	5/34	20/69	25/103	24.2
10,000 *r*.	0/59	1/73	1/132	0.7
50,000 *r*.	60/69	45/45	105/114	94.9

TABLE II

EXPERIMENTAL DATA: EMBRYOS AT 14°C.

Irradiation of Sperm Roentgen Units	No. Hatched	No. Inseminated		Total	Percentage Hatched
Control.	67/69	48/48	45/45	160/162	98.8%
15 *r*.	67/70	47/47	65/65	179/182	98.4
25 *r*.	49/58	58/58	60/60	167/176	94.8
50 *r*.	50/66	56/56	51/55	157/177	88.7
100 *r*.	50/82	64/64	56/62	170/208	86.5
250 *r*. . . .	37/79	30/59	40/47	107/185	58.0
500 *r*.	12/68	5/76	52/72	69/216	31.9
1,000 *r*.	6/64	0/79	40/80	46/223	20.6
10,000 *r*. . . .	0/60	1/71	4/77	5/208	2.4
50,000 *r*. . . .	80/100	65/77	41/41	186/218	86.1

the ultimate expression of the induced lethal condition. If we then average the two above tables we find that in all cases but one there is a total of more than 300 embryos considered at each of the radiation levels (Table III).

TABLE III

HATCHING DATA

X-Radiation Dose	At 14°C.	At 23°C.	Average	Total Embryos
Control. . .	98.8%	98.3%	98.5%	332
15 *r*.	98.4	97.3	97.8	367
25 *r*.	94.8	96.1	95.5	333
50 *r*.	88.7	94.3	91.9	282
100 *r*.	86.5	83.6	88.0	306
250 *r*. . . .	58.0	78.6	68.3	325
500 *r*.	31.9	55.4	43.7	308
1,000 *r*.	20.6	24.2	22.4	326
10,000 *r*.	2.4	0.7	1.6	340
50,000 *r*.	86.1	94.9	90.5	332

In each experimental series a second control group was secured at the end of the experiment, using the same material as in the earlier control group. This was simply a means of checking the normal viability of the sperm and eggs after the period of radiations. The results were almost identical with the initial control group, hence are not included in the above data. The composite data of Table III is plotted in Fig. 1 below.

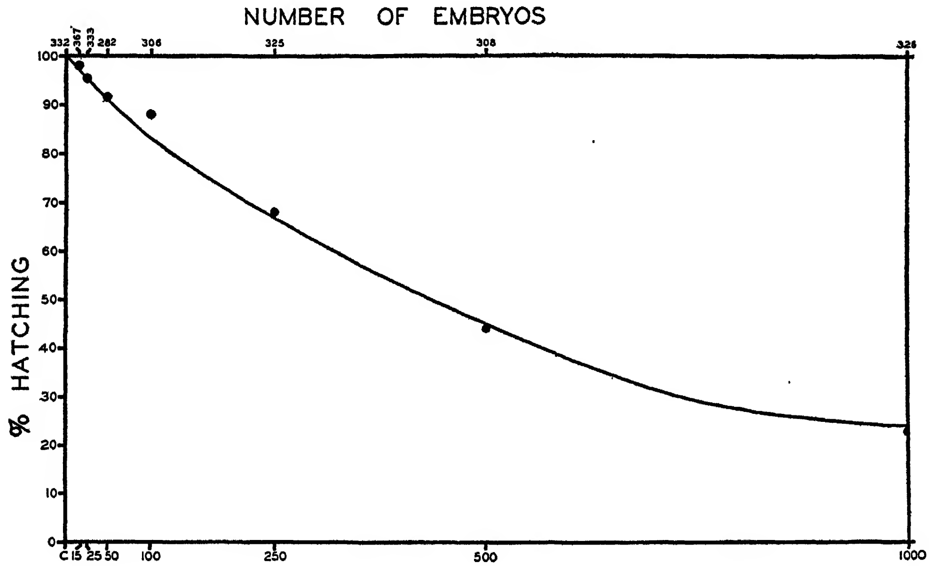


Fig. 1. X-ray exposure: roentgens. Frog sperm x-rayed before insemination. Hatching data.

In order to test the statement of Hertwig that with increasing x-ray exposure there was increase in tendency toward normal development, the range between 10,000 *r* and 50,000 *r* was broken up into convenient intervals. The hatching data at these high levels are summarized in Table IV and in Fig. 2.

TABLE IV

Hatching Data

Radiation of Sperm	No. Hatching/No. Inseminated			Total	Percentage Hatched
10,000 *r*.	12/340 (data above)			12/340	1.6%
20,000 *r*.	1/75	3/68	2/59	6/202	3.0
30,000 *r*. . . .	8/66	13/82	8/47	29/195	14.9
40,000 *r*. . . .	23/72	46/93	20/64	89/229	38.8
50,000 *r*.	110/332 (data above)			110/332	90.5
Control. . . .	318/321			318/321	98.9

The scale of Fig. 1 must not be confused with that of Fig. 2 as the radiation doses in this latter case are much larger. The rise of the curve from a point somewhere between 10,000 r and 20,000 r is significant. Exposure was carried as high as 65,000 r, but the hatching percentage never quite reached the control level although it attained a point above 90 per cent where the spermatozoa had been exposed to very high x-ray doses. Pl. IV shows various stages of development of embryos whose sperm were exposed to 10,000 r; 50,000 r and the controls. A study of

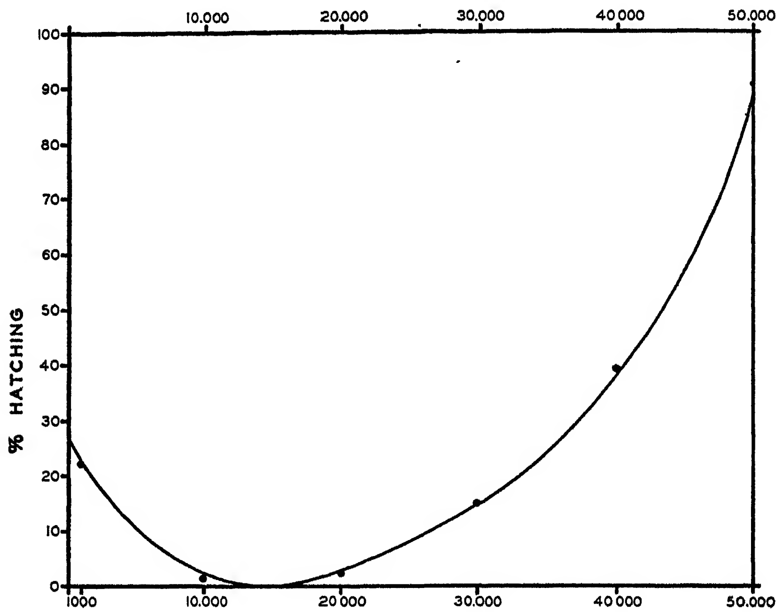


FIG. 2. X-ray exposure: roentgens. Frog sperm x-rayed before insemination. Hatching data.

these photographs will indicate certain rather consistent features of the 50,000 r embryos as compared with the controls. There is a dorso-ventral widening of the tail; a telescoping of the body; a slight reduction of the face, simulating microcephaly; and accompanying reduction of gill size. Cases of exo-gastrulation and spina bifida rarely occurred here as compared with the lower x-ray doses but there seemed to be considerably more œdema (see Pls. III and IV). The embryos from sperm radiated with 50,000 r are comparable to those Hertwig described as parthenogenetic but should more properly be called gynogenetic. There is some evidence, to be investigated and reported in detail later, that these embryos are truly haploid.

An entire experimental series was allowed to develop as far as possible, at laboratory temperatures, until the tadpoles had hatched

and had had several days of activity. This period represented approximately ten days so that all embryos that could hatch had certainly emerged from their jelly. These groups of embryos were then photographed so as to indicate graphically the effect of increasing x-ray exposure of the spermatozoa on the ultimate life of the frog embryo. Pl. V shows the series from the controls through 1,000 *r* and Pl. VI shows the series from 1,000 *r* through 50,000 *r*. It will be noted that the most drastic effects on the embryos occurred at 500 *r* which is considered the sterilizing dose for human beings. It is impossible to compare amphibian and mammalian material except to suggest that frog sperm may be somewhat more resistant to x-radiations than are mammalian sperm. Attention is called to the uniformity of the tadpoles at the 50,000 *r* level as compared with any other groups above the 250 *r* level. This uniformity is in line with the suggestion of parthenogenetic (or gynogenetic) development.

Discussion

The physical conditions utilized in this paper are easily reproducible. Modern x-ray equipment is readily available and at the higher frequencies, such as the 200 kv. used here, the dosage measurements are highly accurate. It is possible to calculate quite accurately the amount of energy absorbed by a stated small mass of biological material. The field of exposure (6.8 cm. diameter, 3 mm. depth at 32.5 cm.) was made the more uniform by the fact that the actively swimming spermatozoa became their own stirring mechanism. Each sperm sample was removed after first stirring, with pipette, the entire suspension.

The biological material was exceedingly uniform. The sperm of *Rana pipiens* are all mature in the fall and dissection of the testis of the hibernating frog provides a very uniform suspension of functional gametes. There is relatively very little interstitial tissue in the frog testis. Sperm, unlike the eggs of the same species, do not need to go through any maturation divisions but are ready for artificial insemination immediately upon removal from the testis. There is no experimental evidence to suggest any difference in the sperm from the two testes of a single male, nor in testes from different hibernating and mature males. In order to avoid any possible experimental error in this direction, however, it was the practice to make all suspensions from a number of males and use the second testis from each of the animals to make up the control suspension.

With the physical and the biological aspects of the problem thoroughly controlled, the sources of error were considerably reduced. Such

error as might exist in this type of experimentation would relate to (1) subjective analysis of the morphogenetic abnormalities and (2) culture conditions under which the embryos are kept. The first source of error is reduced by using the photographic method of presentation of the data. It was found to be very difficult to catalogue abnormalities ranging from exo-gastrulation to hemi-embryos to spina bifida and to draw the lines of distinction sharply, for these are developmental abnormalities which might well occur in sequence. In regard to the culture conditions, the only variation was in the number of embryos per unit of volume. In the regulation finger-bowls used it has been determined that a maximum of about 100 isolated embryos can develop normally. In the majority of cases in this series from 40 to 80 embryos were placed in a single dish. The effect of crowding, when below the maximum for normal development, does not manifest itself except in the size of the ultimate tadpole after a period of weeks (Rugh '34). In most cases the control number exceeded any of the experimental groups, with perfectly normal development.

A clear cut distinction is drawn between the quantitative and the qualitative data. The quantitative data relate to the arbitrarily chosen critical end-point, *i.e.*, hatching. Many embryos die after hatching and some that do not reach the hatching stage die as gastrulæ. This end-point is the most satisfactory one available when we are dealing with such large x-ray doses. With the lower exposures it would be well to carry the results to a second generation where genotypic variations might become manifest. Hatching represents a certain major achievement on the part of the embryo, a certain developmental attainment even though death may ensue. When hatching is thus considered, the photographic series (Pls. V and VI) and Figs. 1 and 2 show graphically the effect of x-radiation of frog spermatozoa.

Cytological studies on qualitative effects are contemplated, but the macroscopic analysis may be presented here. The great variety of abnormalities are indicative of the variations in damage to different sperm cells. But the same abnormalities have been induced by exposure of the *developing embryo* of the frog to supra-maximal temperatures for short periods (Hoadley '38) and also to sub-lethal doses of 2–4 dinitrophenol (Dawson '38). Sudden and drastic reduction in temperature will also induce similar abnormalities so that we are not dealing with factors peculiar to x-ray effects. The major difference between the data of this paper and these other observations is that here the abnormalities are produced by subjecting only half of the zygote, the

sperm, to x-rays prior to syngamy. Since the sperm is merely a naked nucleus, the embryonic disruptions are caused by some change within the sperm nucleus, caused by the x-rays.

Such effects on the sperm nucleus cannot be induced by supramaximal temperatures. Sperm suspensions subjected to lethal temperatures of 32°C. or higher for three minutes are unaffected but by four minutes all of the sperm are killed. Embryos resulting from sperm subjected to this temperature for only three minutes are perfectly normal and do not show any of the x-ray induced abnormalities.

While the x-ray effects on the sperm become manifest with development of the embryo, there is no effect on the motility or the fertilizing power of the sperm itself at any x-ray dosage. Early development of the embryo is unaffected (*i.e.*, cleavage and blastula stages). But this is not surprising for many amphibian hybrid crosses will go to gastrulation and then die, indicating that early development is gynogenic. That the sperm shows no immediate morphological change, and in no way affects early development, is in line with the thesis of this paper.

There are at least two major theories regarding the biological effect of radiations (Goodspeed and Uber '39). One theory suggests that there is a mass effect of the x-rays which would, for instance, involve the entire sperm cell. The gamete is damaged to a greater or lesser degree, but it may still be functional. Such a view cannot readily be disproved. In fact one might argue that the damaged sperm nucleus, when fused with the egg nucleus, weakens the zygotic combination so that the normally inconsequential variations of the environment attain major importance. In this way the great variety of abnormalities could be explained. More frequently, however, the explanation is given in terms of varying inherent radio-sensitivity of different (sperm) cells.

The second theory, known as the Quantum Hit Theory, suggests that the effective ionization (excitation) is limited to certain very small sensitive volumes in the protoplasmic system, generally assumed to be within the chromosome or the gene. To be effective, quanta must be absorbed by the sensitive volume and the variety of results is said to be due to the discontinuous nature of radiation quanta absorption. Survival curves (hatching curve) if plotted as a function of dosage should be exponential. This is essentially the case in Figs. 1 and 2, suggesting the effectiveness of one or more quanta.

This second theory has the support of both plant and animal geneticists. Muller ('27) was one of the first to transmute genes artificially with x-rays, followed shortly by Timofeeff-Ressovsky ('29). In

many cases there are visible changes in the chromosomes (Muller and Altenburg '28 and '30). However, with genic maps accurately determined, transgenations were demonstrated which included deletion (Muller '35; Demerec and Hoover '36); or duplication (Bridges '36; Muller '36; Muller, Prokofyeva-Belgovskaya and Kossikov '36) of such small chromosomal sections as to incorporate only one or two genes. Individual genes have been shuffled about by x-rays (Muller and Prokofyeva '35; Bridges '36; Dobzhansky '36; Grunberg '37). It has not been determined whether the x-ray effect may be due to a precipitation of protein (Needham '31; Clark '36) or to some chemical change within the molecule (Patterson and Muller '30). Goldschmidt ('37), in fact, does away with such discrete units as genes and ascribes to the whole chromosome an architecture responsible for all the so-called genic manifestations. The normal architecture is needed for normal development.

Using the word mutation in its broadest sense (heritable change) to include any chromosomal re-arrangement, whether it be whole chromosomes, fragments, or genes, there is abundant evidence to show that lethal mutations can be induced by x-rays. Most recently this has been shown in maize (Stadler '32, '36) and in the parasitic wasp, Habrobracon (Whiting '37, '38; Bishop '37; Maxwell '38). Whiting demonstrated that x-ray treatment of Habrobracon sperm with doses of 20,000 *r* or more produces at least one dominant lethal in every sperm cell. He found it much easier to induce dominant lethals than to kill or inactivate the sperm by x-raying.

All of the above data on the effect on the embryo by x-radiation of frog sperm can best be explained on the basis of x-ray induced (lethal) mutation, effected by the random bombardment of extremely minute (but sensitive) portions of the sperm chromosomes. At the low dosages relatively few spermatozoa are hit at all and some of these may be hit in inconsequential regions. As the bombardment continues, more sperm are hit at points which are important for the normal development of the post-gastrulation embryo, until at 500 *r* there is morphological evidence that practically every sperm in the suspension has been hit at least once. No doubt many of the sperm exposed to even lower doses are hit a number of times, and if a single hit, or the summation of several hits, is sufficiently important from a morphogenetic point of view the embryo will manifest gross abnormalities which may prove lethal. As the number of hits increases (x-ray dose increase) there comes a time when every sperm nucleus in the suspension is sufficiently damaged to render

it useless in syngamy (20,000 r) and all that remains for the sperm to do is the activation of the egg. This we find to be the case above 20,000 r where there is evidence of gynogenetic development.

Experimental evidence along cytological lines will be presented in a future paper. Suffice it to say that some preliminary studies on the chromosomal complexes of embryos, destined to die as a result of x-radiation of the sperm, have indicated support of this explanation of x-ray induced lethal conditions. Final verification will come through tissue culture and transplantation studies which are in progress.

Finally, we are dealing in this type of study with the fundamental problem of growth, the problem of organized development. This is a dynamic process and cannot be compared with the effect of x-radiation upon a protein or an electrical field. A delicate and yet exceedingly important change is brought about in a sperm cell before it is allowed to initiate development in the normal egg and the effect of this change (mutation) is not manifested immediately but during synthesis, the organization and the integration of the parts of the vital organism. The organism is made up of innumerable but essentially uniform cellular units, and if these units are disrupted or if certain of them acquire special importance (organizer field) and these are unable to go through normal mitosis, then there appears a situation which the organism as a whole cannot control or bring into line with normal development. It is the contention of this paper, supported thus far by macroscopic analysis, that there may be whole chromosomes, genes, molecules, or even atoms that are specifically necessary for normal organogeney and that even at the lowest x-ray exposures some of these vital units may be seriously damaged. When x-ray damage of sperm is sufficient to cause the ultimate death of the embryo, we are dealing with a lethal mutation whether it involves an ultra-microscopic atom or a whole chromosome.

Summary

1. Frog sperm, in suspension, were irradiated with x-rays at 200 kv. with doses ranging from 15 r to 50,000 r prior to insemination of normal eggs.

2. X-radiation of frog sperm had no effect on their motility nor on their powers of fertilization.

3. No irradiated sperm, at any x-ray dosage, had any effect on the rate or the pattern of cleavage, prior to gastrulation.

4. The earliest manifestation of any abnormality, induced by the x-radiation of frog sperm, occurred at gastrulation. Such abnormalities

included exo-gastrulation, and, at later stages, hemi-embryos and spina bifida.

5. As the x-ray exposure of sperm increased from 15 *r* to 10,000 *r* the percentage of abnormal embryos increased and the lethal end-point of development was brought back to a point nearer the gastrula stage, but never earlier than the gastrula. At 250 *r* 68 per cent hatched and many died after hatching but at 10,000 *r* only 1.6 per cent hatched and the majority died as gastrulæ or neurulæ.

6. As the x-ray exposure of sperm was increased above 10,000 *r* there was a rising curve of hatching percentage which reached 90.5 per cent at 50,000 *r*. These embryos were morphologically quite uniform, and very similar to haploids otherwise produced. They showed dorsal flexure; dorso-ventral widening of the tail; telescoping of the body; and some microcephaly. There is evidence that sperm subjected to these high doses are incapacitated for syngamy but are still able to activate the egg to gynogenetic development.

7. Alterations in embryogeny resulting from x-radiation of spermatozoa manifest themselves at or after gastrulation. The disruptive factors may, however, be active before this time. Exo-gastrulation and spina bifida are indications of irregular cell movements at gastrulation.

8. There is no uniformity in the type of abnormality produced in the embryo by x-radiation of frog sperm. Similar abnormalities may be produced by subjecting the developing embryo to certain physical or chemical agents. In these experiments the physical agent was the x-ray applied to half the zygote, the sperm.

9. A tentative explanation is offered for the x-ray effects produced, based somewhat upon parallel experiments and results with both plant and animal material. It is suggested that both dominant and recessive (lethal) mutations may be produced by x-ray bombardment of frog sperm chromosomes. Such an explanation is in line with the fact that some sperm are severely damaged even at 15 *r* and at 1,000 *r* some sperm may escape bombardment altogether. At 10,000 *r* it is suggested that all sperm carry at least one dominant lethal factor. The wide variety of embryonic effects at the lower doses is similarly explained. The term mutation is here used in its broadest sense.

10. At the lowest exposure some spermatozoa are severely damaged, while at the higher dosages all of the exposed spermatozoa are hit. Death prior to metamorphosis may be considered as resulting from induced embryonic (dominant) lethal mutations. These mutations never kill the embryo before gastrulation. The exact time of death

depends upon the number and the importance (from the morphogenetic point of view) of chromosomal (genic) changes produced by x-ray bombardment.

11. These observations were made primarily as a basis for a series of planned experiments which will include the study of the wave-length effects; recovery; cytological effects of x-rays correlated with embryonic development; and survival of irradiated tissue in culture and in transplantations to normal host environments.

LITERATURE CITED

Ancel, P. and Vintemberger, P. 1925. Influence de la rapidité d'evolution de l'oeuf irradié de Rana sur la duree de sa survié. *Compt. rendu Soc. de Biol.*, **93**, 281.

——. 1927. Influence de la temperature sur la radiosensibilité d'oeufs d'Oiseaux et de Batraciens. *Compt. rendu Soc. de Biol.*, **97**, 796.

Baldwin, W. M. 1919. The Artificial Production of Monsters Conforming to a Definite Type by Means of X-rays. *Anat. Rec.*, **17**, 135.

——. 1920. The Artificial Production of Monsters Demonstrating Localized Defects as the Result of Injury from X-rays. *Am. Jour. Physiol.*, 52, 296.

——. 1921. The Artificial Production of Syringomyelocœle in the Tadpole by Means of X-rays. *Anat. Rec.*, 22, 305.

Bardeen, C. R. 1907. Abnormal Development of Toad Ova Fertilized by Spermatozoa Exposed to Rœntgen Rays. *Jour. Exp. Zool.*, **4**, 1.

——. 1909. Variations in Susceptibility of Amphibian Ova to the X-rays at Different Stages of Development. *Anat. Rec.*, **3**, 163.

——. 1911. Further Studies on the Variation in Susceptibility of Amphibian Ova to X-rays at Different Stages of Development. *Am. Jour. Anat.*, **11**, 419.

Bishop, D. W. 1937. Induction of Dominant Lethal Effects of X-radiation in Habrobracon. *Genetics*, **22**, 452.

Bohn, G. 1903. Influence des rayons du radium sur les oeufs vierges et fecondes, et sur les premiers stades du developpement. *Compt. rendu Soc. de Biol.*, **136**, 1085.

Bridges, C. B. 1936. The Bar 'gene' a Duplication. *Science*, **83**, 210.

Clark, J. H. 1936. Biological Effects of Radiation. Chap. VIII (McGraw-Hill, New York).

Curtis, W. C., J. A. Cameron, K. O. Mills. 1936. Exogastrulation in Amphibia after x-ray exposure. *Science*, **83**, 354.

Dawson, A. B. 1938. Effects of 2,4 Dinitrophenol on the Early Development of the Frog, *Rana pipiens*. *Jour. Exp. Zool.*, **78**, 101.

Demerec, M. and Hoover, M. E. 1936. Three Related X-chromosome Deficiencies in Drosophila Differing in Extent of Deleted Material, and in Viability. *Jour. Heredity*, **27**, 207.

Dobzhansky, T. 1936. Position Effects on Genes. *Biol. Rev.*, **11**, 364.

Gager, C. S. 1908. Effects of the Rays of Radium on Plants. *Mem. N. Y. Botanical Garden*, 4.

Gilman, P. K. and Baetjer, F. H. 1904. Some Effects of the Rœntgen Rays on the Development of Embryos. *Am. Jour. Physiol.*, **10**, 222.

Goldschmidt, R. 1937. Spontaneous Chromatin Rearrangements and the Theory of the Gene. *Proc. Nat. Acad. Sci.*, **23**, 621.

Goodspeed, T. H. and Uber, F. M. 1939. Radiation and Plant Cytogenetics. *Botanical Rev.*, **5**, 1.

Grünberg, H. 1937. The Position Effect Proved by a Spontaneous Reinversion of the X-chromosome in *Drosophila melanogaster*. *Jour. Genetics*, **34**, 169.

Hertwig, G. 1911. Radiumbestrahlung unbefruchteter Froscheier und ihre Entwicklung nach Befruchtung mit normalem Samen. *Arch. mikrosk. Anat.*, **77**, 165.

Hertwig, O. 1910. Die Radiumstrahlung in ihrer Wirkung auf die Entwicklung tierischer Eier. *Sitzungsb. K. Pruss. Akad. Wissensch.*, 24 Feb. & 28 Juli.

——. 1911. Die Radiumkrankheit tierischer Keimzellen. Ein Beitrag zur experimentellen Zeugungs- und Vererbungslehre. *Arch. mikrosk. Anat.*, **77**, 1 and **77**, 97.

Hertwig, P. 1911. Durch Radiumbestrahlung hervorgerufene Veränderungen in den Kernteilungsfiguren der Eier von Ascaris megalocephala. *Arch. mikrosk. Anat.*, **77**, 301.
Hertwig, Paula. 1916. Durch Radiumbestrahlung verursachte Entwicklung von halbkernigen Triton- und Fischembryonen. *Arch. mikrosk. Anat.*, **87**, 63.
Hoadley, L. 1938. The Effects of Supramaximum Temperatures on the Development of *Rana pipiens*. *Growth*, **2**, 25.
Hoffman, V. 1922. Über Erregung und Lahmung tierischer Zellen durch Röntgenstrahlen. I. Experimentelle Untersuchungen an Froscheiern und Larven. *Strahlentherapie*, **13**, 285.
Holtfreter, J. 1931. Über die Aufzucht isolierter Teile des Amphibienkeimes. *Arch. f. Ent. Mech. der Org.*, **124**, 404.
Levy, O. 1906. Mikroscopische Untersuchungen zu Experimenten über den Einfluss der Radiumstrahlen auf embryonale und regenerative Entwicklung. Nach den hinterlassenen Praparation von Prof. Alfred Schaper. *Roux. Arch. Entwicklungsmech. der Org.*, **21**, 130.
Maxwell, J. 1938. Inactivation of Sperm by X-radiation in Habrobracon. *Biol. Bull.*, **74**, 253.
McGregor, J. H. 1908. Abnormal Development of Frog Embryos as a Result of Treatment of Ova and Sperm with Rœntgen Rays. *Science*, **27**, 445.
Muller, H. J. 1927. Artificial Transmutation of the Gene. *Science*, **66**, 84.
——. 1935. A Viable Two-gene Deficiency Phenotypically Resembling the Corresponding Hypomorphic Mutations. *Jour. Hered.*, **26**, 469.
——. 1936. Bar Duplication. *Science*, **83**, 528.
Muller, H. J. and Altenburg, E. 1928. Chromosome Translocations Produced by X-rays in Drosophila. *Anat. Rev.*, **41**, 100.
——. 1930. The Frequency of Translocations Produced by X-rays in Drosophila. *Genetics*, **15**, 283.
Muller, H. J. and Prokofyeva, A. A. 1935. The Individual Gene in Relation to the Chromomere and the Chromosome. *Proc. Nat. Acad. Sci.*, **21**, 16.
Muller, H. J., Prokofyeva, A. A., and Kossikov, K. V. 1936. Unequal Crossing-over in the Bar Mutant as a Result of Duplication of a Minute Chromosome of Drosophila. *Compt. rendu. Acad. Sci. USSR*, **2**, 78.
Needham, J. 1931. Chemical Embryology. Vol. III, p. 1435. Cambridge.
Olivieri, F. 1929. Dell azione dei raggi Roentgen sulle di Bufo vulgaris. *Monit. Zool. Ital.*, **40** (11/12), 442–444. *Riv. Radiol. e Fisica Med.*, **1**, 527.
Packard, C. 1914. The Effects of Radium Radiations on the Fertilization of Nereis. *Jour. Exp. Zool.*, **16**.
Patterson, J. T. and Muller, H. J. 1930. Are "Progressive" Mutations Produced by X-rays? *Genetics*, **15**, 495.
Richards, A. 1914. The Effects of X-rays on the Action of Certain Enzymes. *Am. Jour. Physiol.*, **35**, 224.
——. 1915. Recent Studies on the Biological Effects of Radioactivity. *Science*, **42**, 287.
Richards, A. and Good, D. J. 1919. Notes on the Effect of X-radiation on the Development of Cumingia Eggs. *Biol. Bull.*, **37**, 209.
Richards, A. and Woodward, A. E. 1915. Note on the Effect of X-radiation on Fertilizin. *Biol. Bull.*, **28**, 140.
Rugh, R. 1934. The Space Factor in the Growth Rate of Tadpoles. *Ecology*, **15**, 407.
——. 1935. Induced Ovulation and Artificial Fertilization in the Frog. *Biol. Bull.*, **66**, 22.
Schaper, A. 1904. Experimentelle Untersuchungen über den Einfluss der Radium-strahlen und der Radiumemanation auf embryonale und regenerative Entwicklungsborgange. *Anat. Anz.*, **25**, 298, 326.
Schwartz, G. 1903. Über die Wirkung den Radiumstrahlen; eine physiologische chemische Studie am Huhnerei. *Arch. f. Physiol.*, **100**, 532.
Stadler, L. J. 1932. On the Genetic Nature of Induced Mutations in Plants. *Proc. 6th Int. Congress of Genetics*, **1**, 274.
Stadler, L. J. and Sprague, G. F. 1936. Genetic Effects of Ultra-violet Radiation in Maize. III. Effects of Nearly Monochromatic λ 2537, and Comparison of Effects of X-ray and Ultra-violet Treatment. *Proc. Nat. Acad. Sc.*, **22**, 584.

Timofeeff-Ressovsky, N. W. 1929. The Effect of X-rays in Producing Somatic Genovariations of a Definite Locus in Different Directions in *Drosophila melanogaster*. *Am. Nat.*, **63**, 118.

Whiting, P. W. 1937. Habrobracon as a Means for Testing the Effectiveness of Physical Agents in Causing Mutations. *Proc. Penn. Acad. Sc.*, **11**, 50.

——. 1938. The Induction of Dominant and Recessive Lethals by Radiation in Habrobracon. *Genetics*, **23**, 562.

Wolf, O. M. 1929. Effect of Daily Transplants of the Anterior Lobe of the Pituitary on Reproduction of the Frog. *Anat. Rec.*, **44**, 206.

Woskressensky, N. M. 1928. Ueber die Wirkung der Röntgenbestrahlung auf das embryonale Wachstum. *Arch. Entwick. Organ.*, **113**, 447.

EXPLANATION OF PLATES

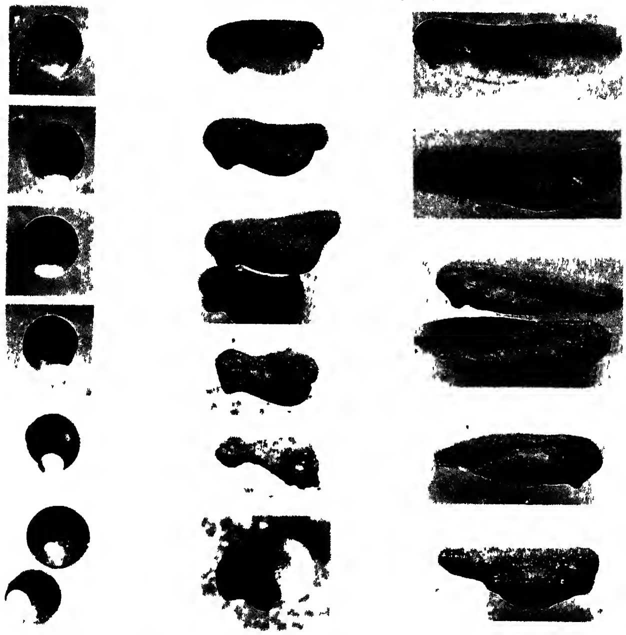
Pl. I. Across the top of this plate is the Control series from the two cell stage to the tadpole, shown on the same scale as all succeeding photographs. Inasmuch as no experimental effects were manifest before gastrulation, the photographs of experimental material begin with the gastrula.

From left to right the columns represent advancing stages of development. Only the extremes (normal to abnormal) at these various stages are shown, with no attempt to indicate percentage of occurrence. Where two embryos are photographed together, this was to indicate degree of variation.

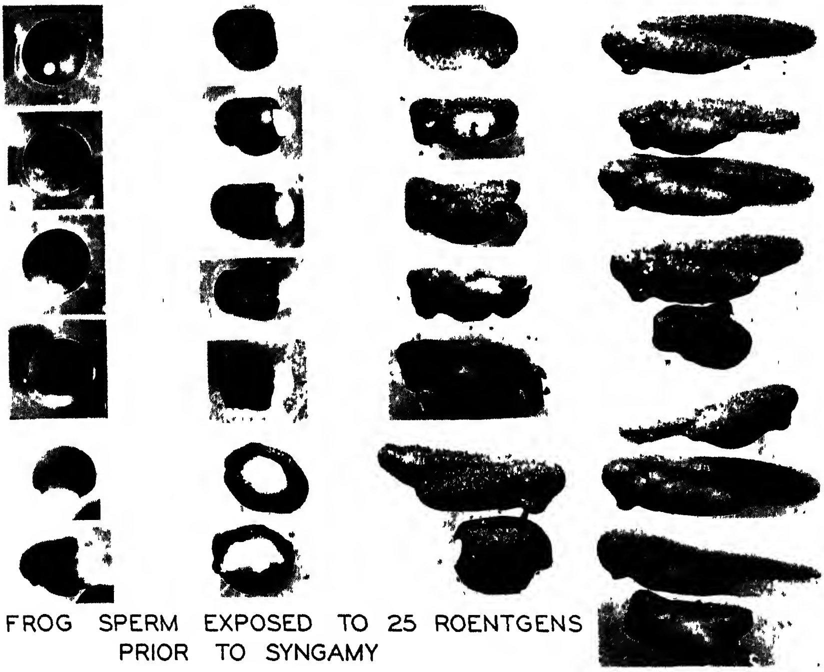
Note both normal and exo-gastrulæ even at 15 *r*; spina bifida and hemi-embryos. This indicates that some spermatozoa were very severely damaged even at this extremely low x-ray exposure. In the same cultures, the majority of embryos developed into perfectly normal tadpoles, phenotypically speaking. The lower series represents those abnormalities seen when the sperm were exposed to 25 *r*.

PLATE I

CONTROL SERIES

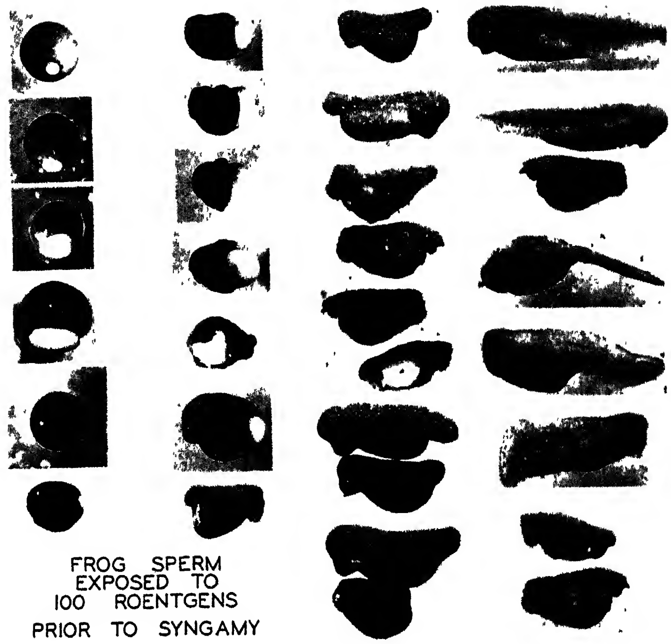


FROG SPERM EXPOSED TO 15 ROENTGENS PRIOR TO SYNGAMY

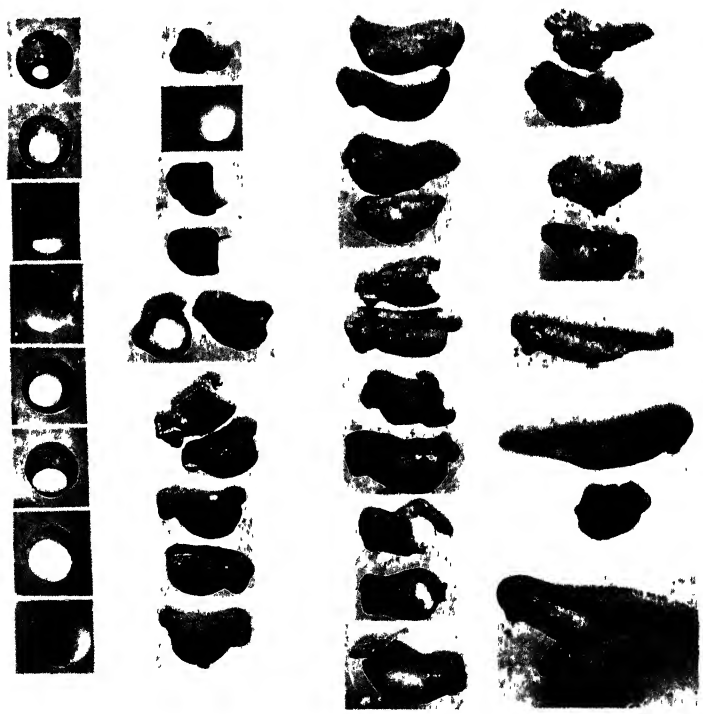


FROG SPERM EXPOSED TO 25 ROENTGENS PRIOR TO SYNGAMY

PLATE II



FROG SPERM
EXPOSED TO
100 ROENTGENS
PRIOR TO SYNGAMY



FROG SPERM EXPOSED TO 500 ROENTGENS
PRIOR TO SYNGAMY

Pl. II. Abnormalities at 100 r exposure of spermatozoa and also at 500 r. At this latter dosage fewer than half of the embryos hatched, but even here there were some apparently normal tadpoles.

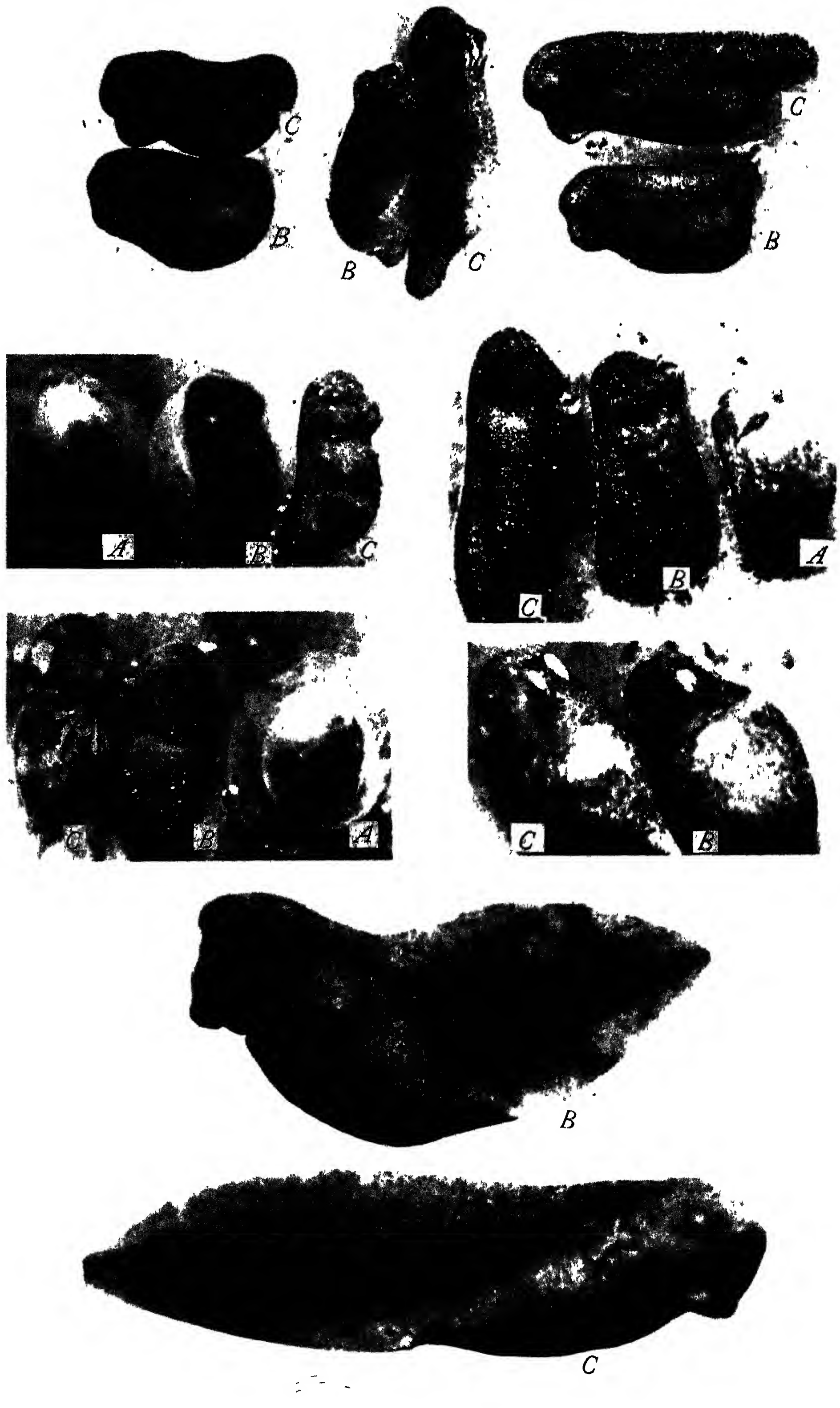
PL. III. Abnormalities induced by exposing spermatozoa to 1,000 *r*; 10,000 *r* and 50,000 *r*. Those at 10,000 *r* at the extreme right did not hatch but were dissected out of their jelly capsules in order to indicate their state of development. The best developed embryos were only amorphous masses of protoplasm, with slight indication of ciliary movement. At 50,000 *r* practically none of the abnormalities common to the preceding groups are manifested. The embryos that hatched are comparable in appearance to those which are produced parthenogentically, by other means. Note appearance of œdema; dorso-ventral widening of the tail; telescoping and flexing of the body; reduction of the gills and tendency toward microcephaly.

FROG SPERM EXPOSED TO 1.000 ROENTGENS
PRIOR TO SYNGAMY

FROG SPERM EXPOSED TO 10.000 ROENTGENS
PRIOR TO SYNGAMY

FROG SPERM EXPOSED TO 50.000 ROENTGENS
PRIOR TO SYNGAMY

PLATE IV



Pl. IV. A comparative series taken at different stages of development to show the differences between the Controls (*C*); those whose sperm received 10,000 *r* exposure (*A*) and those whose sperm received 50,000 *r* exposure (*B*). Below are enlarged (to the same scale) photographs of one of these so-called parthenogenetic (50,000 *r*) embryos compared with the normal control.

PL. V. Entire batches of embryos from sperm receiving increasing exposures to x-radiations. *A*—control; *B*—15 *r*; *C*—25 *r*; *D*—50 *r*; *E*—100 *r*; *F*—250 *r*; *G*—500 *r*; *H*—1,000 *r*. Note that the percentage hatching is progressively decreased with increasing exposure to the sperm to x-rays prior to syngamy. Note also that while abnormalities appear even at 15 *r* the most drastic effects appear at 250 *r* and that at 1,000 *r* a few hatch but all are definitely abnormal.

PLATE V

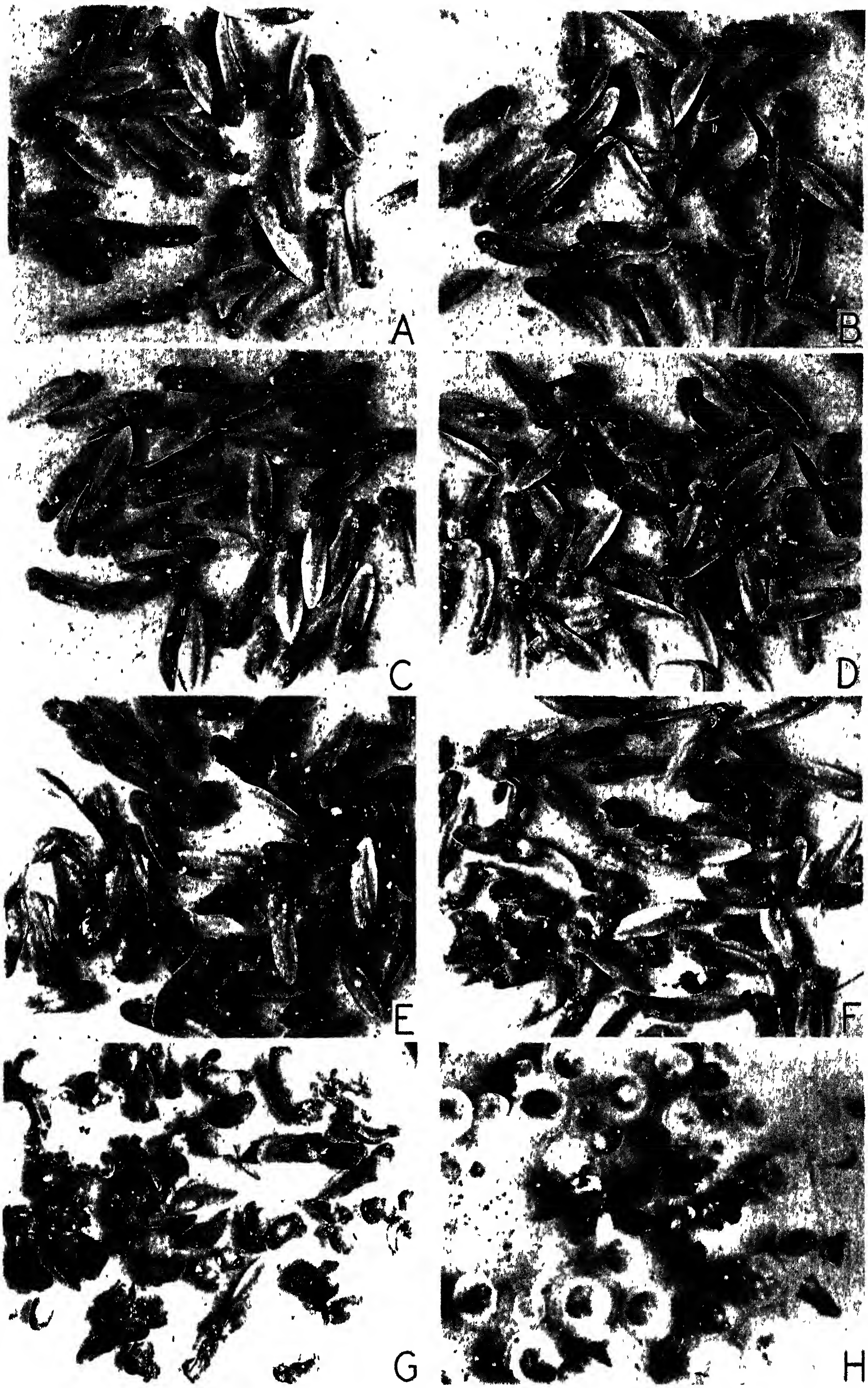
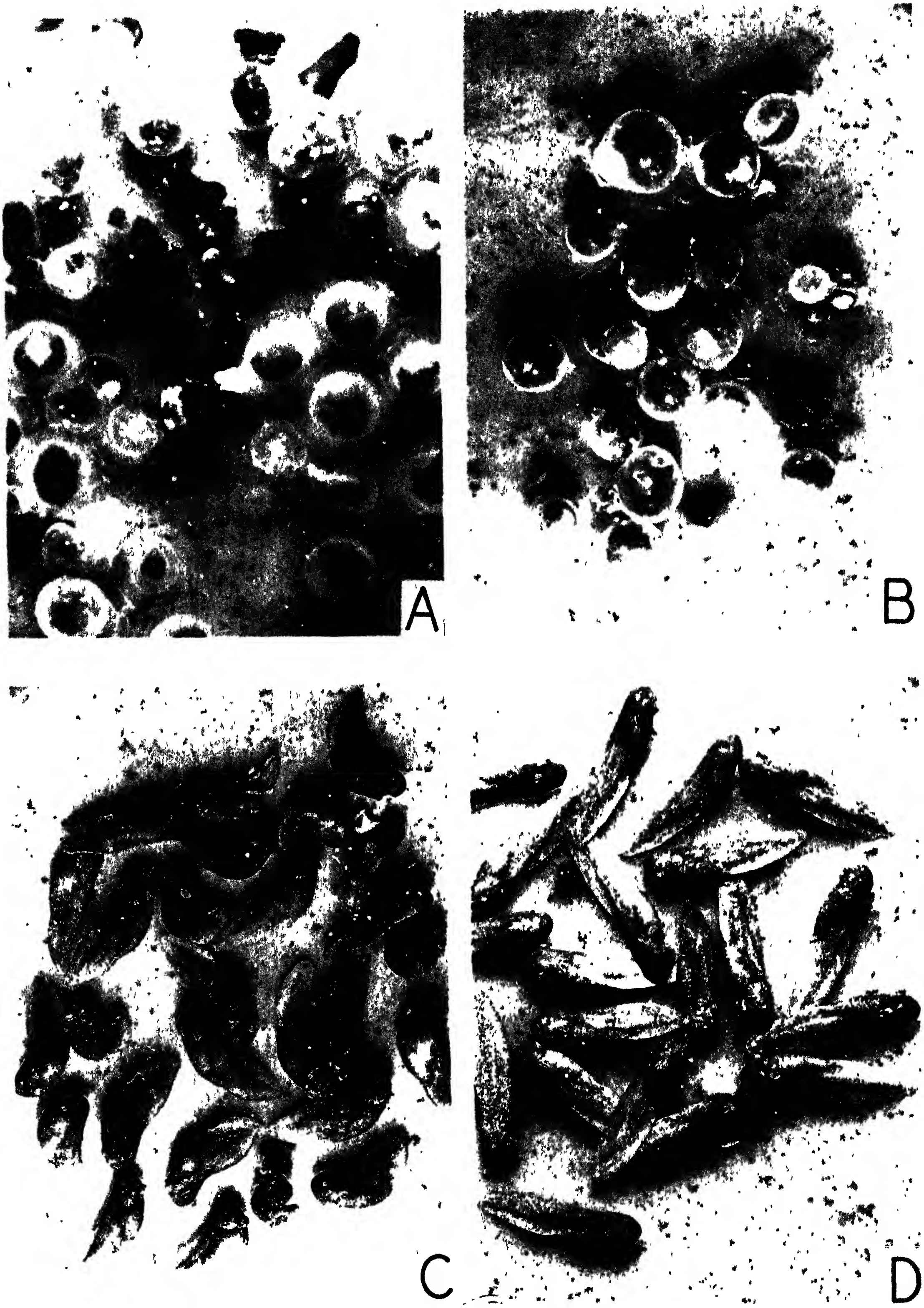


PLATE VI



Pl. VI. Similar series as compared with Pl. V. *A*—1,000 *r*; *B*—10,000 *r*; *C*—50,000 *r*; and *D*—control. Intermediate steps between *B* and *C* were not photographed. Note uniformity of embryos at the 50,000 *r* level. These are regarded as gynogenetically developed.

A STUDY OF FIREBALL RADIANT POSITIONS

FLETCHER WATSON

Harvard College Observatory

(Introduced by Harlow Shapley)

Abstract

The reality of the published hyperbolic velocities for a large portion of bright meteors, which would require their origin in interstellar space, is considered. Their directions of motion, corrected for the effects of the earth's motion, show a definite concentration toward the plane of the solar system. When compared to meteor streams, which almost certainly belong to the system, the bright meteors exhibit an equal if not higher degree of concentration about the ecliptic. Since the study of photographed meteors, which are comparable to fireballs in brightness, has revealed no hyperbolic velocities, and since the directions of motion of fireballs and meteor streams are practically indistinguishable, it seems probable that the velocities previously assigned to the fireballs were too great and that the bodies are permanent members of the solar system.

In the often discussed *Katalog der Bestimmungsgrössen für 611 Bahnen Grosser Meteore* by von Niessl and Hoffmeister (*Denksch. d. Akad. d. Wiss. Wien, Math.-Naturw. Kl.*, **100**, 1925), seventy-nine per cent of the heliocentric velocities are hyperbolic. To Fisher (*H. C.* 331, 1928; *idem*, 375, 1932) this seemed to be an excessive proportion inasmuch as he found a correlation between the positions of the earth in its orbit and the maxima and minima of fireball frequency. Malzev (*M. N.*, **90**, 568, 1929) found that for the same geocentric velocities the bodies with hyperbolic orbits appeared higher and disappeared lower than those with elliptical orbits. Following Fisher's second paper, Malzev demonstrated (*Tashkent Obs. Bull.*, No. 3, p. 57, 1934) that the harmonic mean velocities were superior to those previously derived, but were still principally hyperbolic. Indirect evidence on the velocities was sought by Knopf (*A. N.*, **242**, 161, 1931) who plotted the distribution of the apparent radiants, and indicated ten regions in which the radiants appeared to cluster. When he examined the space radiants of the Taurus and Scorpius streams, Knopf believed he found evidence that these streams were active through the entire year and were truly interstellar. From the study of several well-observed fireballs, Wylie (*P. A.*, **43**, 241, 312, 379, 602, 657, 1935; **44**, 42, 152, 1936; **45**, 101, 209, 1937) concluded that the paths given in the *Katalog* were excessively long, whence both the geocentric and heliocentric velocities had been exaggerated. An upward extension of the path, as suggested by Wylie,

would produce the high altitudes of appearance for high-velocity meteors as found by Malzev. The contradictions which have appeared in the several discussions of the reality of the hyperbolic velocities in the *Katalog* demand that any evidence be presented to aid in clarifying the situation. With the exception of the incomplete reduction by Knopf, none of the discussions have involved the space radiants. It would seem that from a study of the distribution of the space radiants some indications of the reality of the high velocities might be gained.

As a result of the earth's orbital motion, all meteoric radiants are displaced toward the meteoric apex. When a is the elongation of the apparent radiant and V_h is the heliocentric velocity of the body, in kilometers per second, the shift, B, necessary to locate the space radiant, is given by

$$\sin B = 30/V_h \cdot \sin a.$$

The displacement B is along the great circle passing through the apparent radiant and the apex. For the radiants in the *Katalog*, the reductions may be performed with sufficient accuracy on a globe. In the present reduction, I have made use of the velocities given there for 411 fireballs. The distribution derived is shown in Fig. 1, where an appreciable concentration to the ecliptic appears. To separate the bodies with elliptical and hyperbolic orbits, three groups were formed: $V_h < 43$ (dots), $43 < V_h < 50$ (crosses), and $50 < V_h$ (plus signs). The seven radiants from which meteorites fell are encircled. For comparison of the latitude distribution of the three groups, the frequencies have been counted over ten-degree zones north of the ecliptic and reduced to percentages, as given in Table I.

TABLE I

LATITUDE DISTRIBUTION OF SPACE RADIANTS FROM THE VON NIESSL-HOFFMEISTER KATALOG *in percentages*

Mid-Latitude:	5°	15°	25°	35°	45°	55°	65°	75°	85°
Groups									
$V_h < 43$.......	33	20	15	11	10	7	2	2	0
$43 < V_h < 50$..	40	12	20	14	5	5	1	2	1
$50 < V_h$.....	45	18	9	9	9	2	5	2	1

Contrary to expectation, the concentration to the ecliptic increases with the heliocentric velocity. A partial, but probably insufficient, explanation may be found in the fact that but few radiants have been corrected for zenith attraction (*Katalog*, pp. 2 and 28; *Tashkent Obs. Bull.*, No. 3, 57, 1934). The radiants of the "low-velocity" bodies

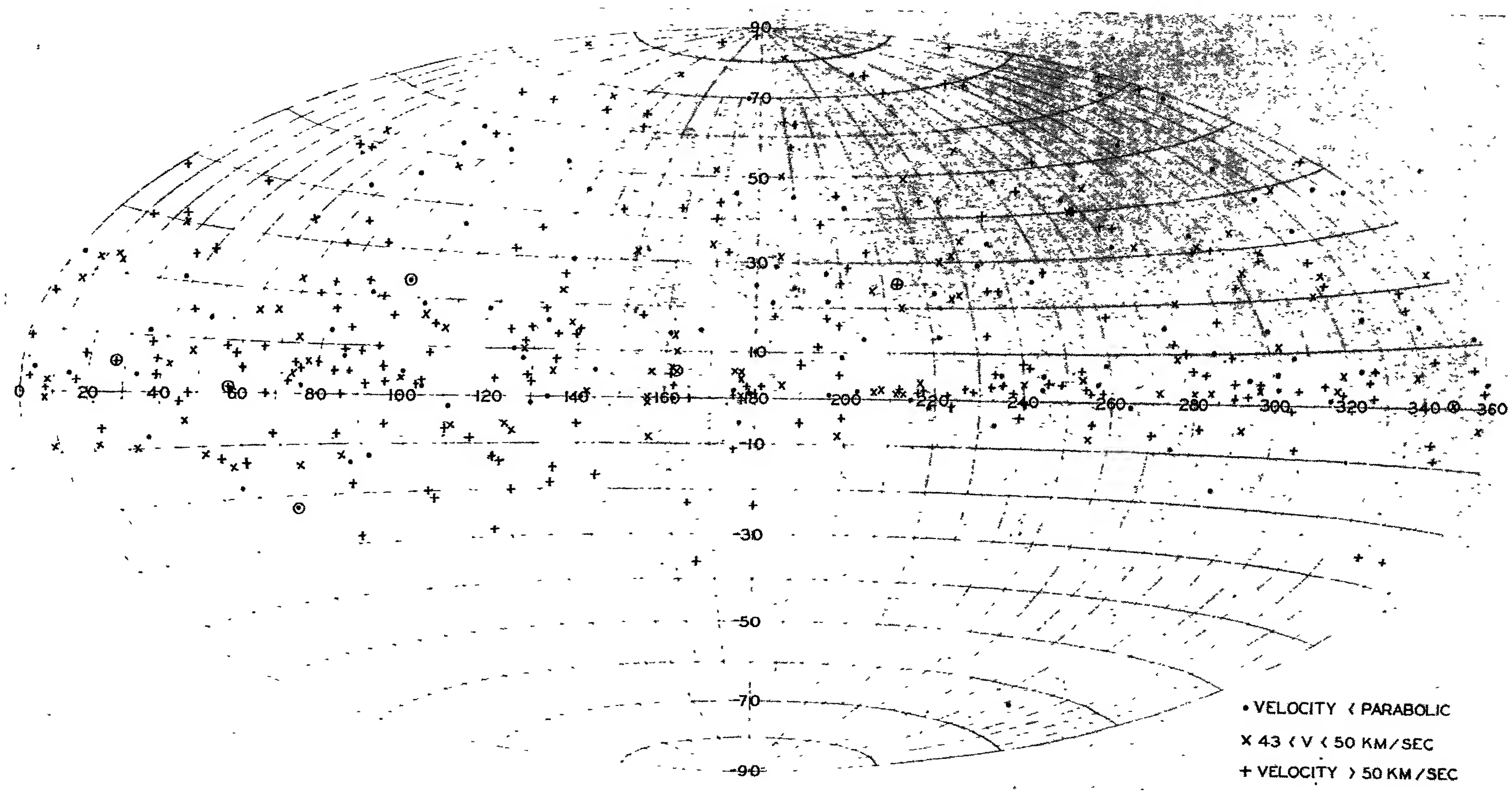


FIG. 1. Distribution of space radiants from the von Niessl-Hoffmeister catalogue.

have been displaced most toward the zenith and, since one component of the displacement is perpendicular to the ecliptic, the ecliptic concentration for bodies of low velocity would be more "washed-out" than for those of high velocity.

Figure 1 shows several groupings of radiants which might indicate streams of fireballs. The most concentrated are near 160°, + 5°; 175°, + 3°; and 210°, + 2°—the Scorpius stream, for which the mean heliocentric velocity is 51 km/sec, with a standard deviation for one velocity of ± 8 km/sec. For the stream members as chosen by von Niessl (*Sitz. d. K. Akad. d. Wiss. Wien*, **121**, p. 1925, 1912), the mean velocity is approximately the same, although the standard deviation of a single velocity is ± 10 km/sec without regard for possible systematic errors. The significance of such groupings may be tested with the Poisson equation for the probable distribution,

$$P(n) = \frac{n_0^n}{n!} e^{-n_0},$$

where n_0 is the average number of radiants per unit area and n is the anticipated number. Between latitudes 0° and + 10° there are 141 radiants which, when considered over 36 intervals ten degrees wide, give $n_0 = 3.91$. The expected and observed distributions are shown in Table II.

TABLE II

n	$P(n)$	$N_{\text{exp.}}$	$N_{\text{obs.}}$	n	$P(n)$	$N_{\text{exp.}}$	$N_{\text{obs.}}$
0	0.024	0.9	2 ± 1.4	5	0.15	5.6	4 ± 2.0
1	.090	3.3	2 1.4	6	.094	3.5	5 2.2
2	.17	6.4	4 2.0	7	.050	1.9	3 1.7
3	.21	7.9	8 2.8	8	.018	0.6	1 1.0
4	.20	7.5	7 2.6				

There are no indications in Table II that the groupings are other than accidental. Either the clustering of radiants observed by Knopf was real and the *Katalog* velocities used to transform them into space radiants were in error, the individual uncertainty being eight to ten kilometers per second; or the apparent clusterings were accidental and the streams spurious. The possibility that the dispersion of the apparent clusterings developed from errors in the velocities does not exclude the possibility that the clusterings were accidental, or influenced by general seasonal effects.

When the material in the *Katalog* is discussed, the distribution of

the local mean times of appearance of the fireballs must be considered. This distribution is given in Table III. Since the maximum rate for photographed meteors occurs near 6^h L.M.T. (Watson, *H. B.* 910, 1939), it is obvious that the *Katalog* does not contain a homogeneous sample of all fireball appearances. Hence any generalizations regarding times of appearance or heights of appearance and disappearance, derived from studies of the paths in the *Katalog*, should be considered with reservations.

TABLE III

LOCAL MEAN TIMES OF APPEARANCE OF FIREBALLS IN THE KATALOG OF VON NIESSL AND HOFFMEISTER

0^h–1^h	16	12^h–13^h	2
1–2	2	13–14	3
2–3	6	14–15	1
3–4	5	15–16	4
4–5	3	16–17	13
5–6	1	17–18	35
6–7	1	18–19	60
7–8	1	19–20	57
8–9	0	20–21	104
9–10	2	21–22	96
10–11	1	22–23	92
11–12	2	23–24	53

For comparison with the data from the *Katalog*, a similar reduction was made of the radiants found by the Arizona Meteor Expedition (*H. C.* 388, 1934). In the reduction parabolic velocities were assumed. The radiant positions listed in *Harvard Circular* 388 were chosen, not as the result of careful measurement and study of the intersections of individual trails, but because there was a general radiation of trails from those regions throughout at least one night. Hence the radiants are probable positions representing the true radiants within a degree or two. The calculation of space radiants need not be accurate, therefore, to more than a degree and the reduction on the sphere is permissible; furthermore, zenith corrections may be neglected. The radiants which could be identified on successive nights or at intervals of a year were counted as a single radiant; questionable identifications were counted separately. Figure 2 shows the resulting distribution. The concentration toward the ecliptic is apparent from Table IV. In the mean, the *Katalog* radiants are more concentrated to the ecliptic than are the radiants of meteor streams, which almost certainly have periodic orbits. We have then a peculiar condition which can be explained on the alternatives that (1) the high-velocity objects from interstellar space

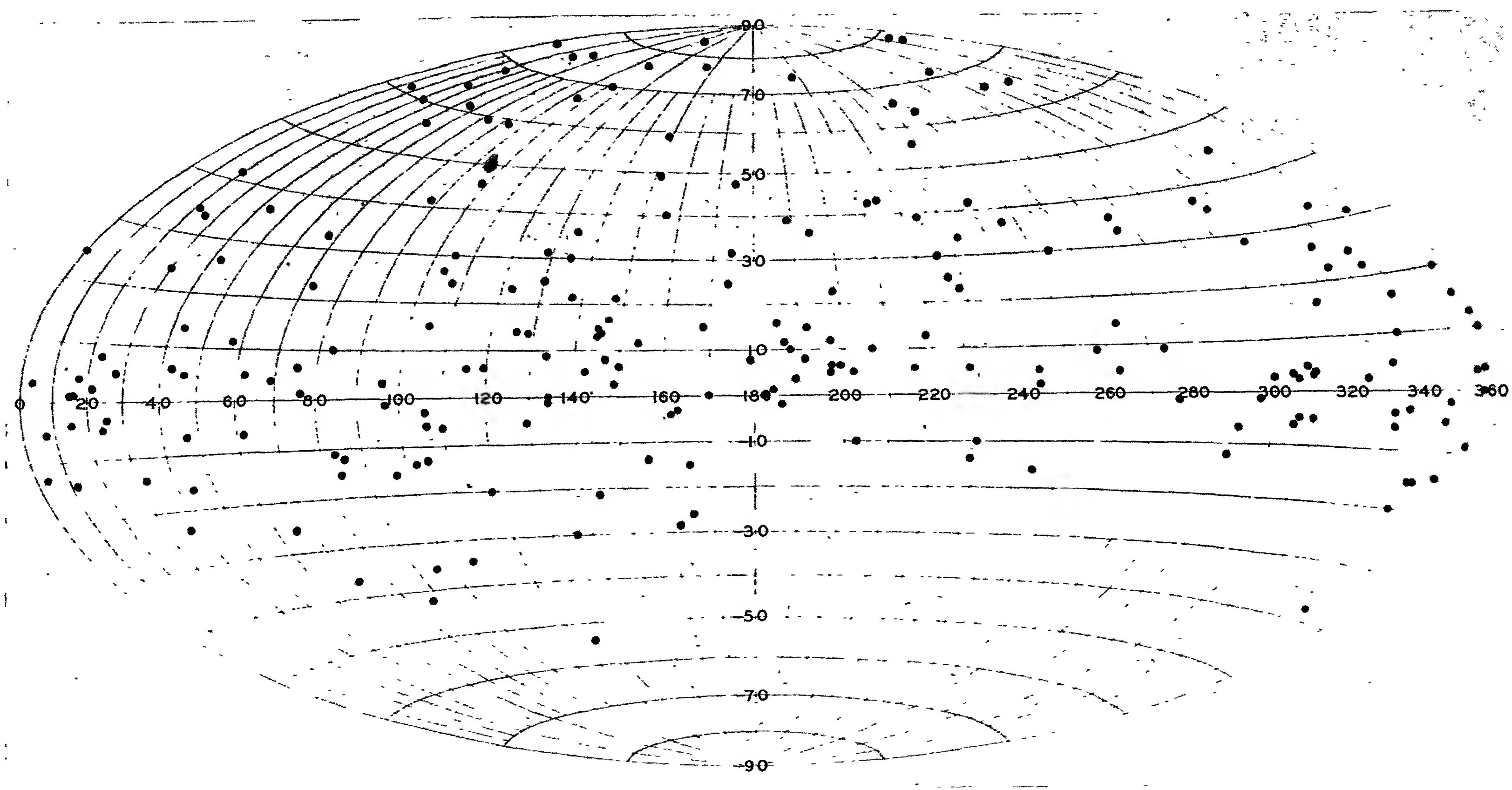


FIG. 2. Distribution of space radiants of meteor showers.

TABLE IV

COMPARISON BETWEEN LATITUDE DISTRIBUTIONS OF KATALOG AND ARIZONA METEOR EXPEDITION SPACE RADIANTS

in percentages

Mid-Latitude	5	15	25	35	45	55	65	75	85
Katalog (mean)	41	17	12	11	8	4	3	2	1
Arizona Exped.	33	14	16	13	6	4	7	6	1

have, by chance, motions principally in the ecliptic, like the periodic streams; or (2) the velocities given in the *Katalog* are not significant and the objects were actually members of the solar system, subject to the perturbing forces which govern the distribution of the orbits of comets and asteroids.

Since the distribution of the space radiants of "high-velocity" fireballs shows no strongly distinguishing characteristics from that of "low-velocity" fireballs and meteor streams, and since Whipple (*Proc. Amer. Phil. Soc.*, 79, 499, 1938; Am. Astr. Soc., meetings of September and December, 1938) has as yet found no evidence for the existence of hyperbolic orbits among photographed meteors, it is probable that the velocities in the von Niessl-Hoffmeister *Katalog* are too great and that the bodies listed there were permanent members of the solar system.

APPARENT DISTRIBUTION OF METEOROID ORBITS

FLETCHER WATSON

Harvard College Observatory

(Introduced by Harlow Shapley)

ABSTRACT

By use of a simple spherical reduction the important elements of parabolic orbits are determined for meteor showers. Despite unfavorable influences from their geocentric velocities, the majority of the streams appear to be moving in the same direction as the earth about the sun, and in orbits not greatly inclined to the plane of the solar system. Streams with perihelion at approximately the earth's distance from the sun are found to be the most frequent.

THE orbital parameters governing the spatial distribution of meteoroids near the sun are principally the inclinations of the orbits and the perihelion distances, which can be computed with ease and speed for parabolic orbits. In statistical groupings over wide ranges such as will be used here, small accidental uncertainties in the individual orbits are unimportant. Thus parabolic orbits are sufficiently accurate, for, even though the true period of revolution about the sun be but a few years, the elements will only infrequently be shifted outside the ranges to be used.

The angle θ between the radius vector and a parabolic orbit at unit distance from the sun is given by the relation

$$\sin^2 \theta = q,$$

where q is the perihelion distance.

From this simple relationship the values of q are quickly found when θ is known. Since all the radius vectors must pass through the sun, the angle θ may be measured on the globe as the arc between the subsolar point and the space radiant, thereby providing directly the value of q.

Any body which collides with the earth must be near a node of its orbit. The inclination of an orbit is then the angle between the ecliptic and the plane containing the radius vector and the vector of heliocentric motion. This angle is most readily measured on the sphere along the great circle 90° from the position of the sun. Thus from the spherical measurements, the space radiants and the important elements

of the parabolic orbits are determined. Small changes in the position of space radiants near the sub-solar point result in large changes in the inclination, and hence the inclinations are least accurate for orbits of small q. Because $q = \sin^2 \theta$, the perihelion distances are influenced most by small changes in θ when θ is near 45°, and the perihelia near 0.5 A.U. are the least accurate.

Although obvious selection factors have influenced the von Niessl-Hoffmeister *Katalog*, the orbits of the fireballs are of interest *per se;* hence the graphical analysis has been utilized for the determination of their parabolic orbits (Table I). The distribution of the inclinations

TABLE I

Distribution of Orbits from von Niessl-Hoffmeister *Katalog*

i \ q	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95 A.U.	Σ
05°	7	9	8	10	11	11	14	12	16	34	132
15	1	2	6	3	3	9	10	12	7	12	65
25	..	4	2	5	5	4	2	10	9	19	60
35	..	2	3	2	2	1	1	4	5	18	38
45	..	1	..	..	3	1	2	3	3	6	19
55	..	1	1	..	1	1	..	2	1	6	13
65	..	..	2	..	..	1	1	2	1	8	15
75	1	..	..	..	..	2	..	1	1	7	12
85	2	1	1	..	1	..	..	1	2	3	11
95	5	..	..	..	..	1	1	..	..	..	7
105	1	..	1	..	..	..	1	..	1	2	6
115	1	2	..	..	..	..	1	..	2	..	6
125	1	..	..	..	..	..	..	1	..	..	2
135	1	..	..	..	..	..	..	..	2	1	4
145	..	..	..	..	..	..	..	..	1	2	3
155	2	..	..	1	1	..	..	..	1	1	6
165	1	1	..	..	..	..	1	1	1	3	8
175	..	..	..	..	1	1	1	1	..	..	4
	23	23	24	21	28	32	35	50	53	122	411

and perihelia near unit distance is in part associated with the distribution of the times of appearance. Even though these data cannot be accepted as representative of fireballs in general, we cannot escape the conclusion that no matter what velocities are taken the great majority of the bodies listed in the *Katalog* were moving in direct orbits of low inclination.

Table II gives for the Arizona radiants, discussed in the preceding paper, the distribution of the apparent elongations, a, from the meteoric apex. Since the radiants from retrograde parabolic orbits must appear within 56° of the apex, the table shows that the majority of the shower-

producing swarms move in direct orbits. In column A, where the frequency of the true elongation is given, we find that nearly two-thirds of the orbits are direct, despite the fact that, because of its increased geocentric velocity, a body moving in a retrograde orbit appears much brighter than one moving in a direct orbit, and showers with retrograde motion are therefore far more conspicuous than those moving directly. In the last two columns of Table II the apparent distributions are reduced to relative number per unit area of the sky, from which the predominance of direct motions is even more strongly evident.

TABLE II

DISTRIBUTION OF APPARENT (a) AND TRUE (A) ELONGATIONS OF ARIZONA RADIANTS FROM THE METEORIC APEX

Elongation	Observed Frequency		Relative Number per Unit Area	
	a	A	a	A
5°	4	2	26.7	13.3
15	8	4	17.6	8.9
25	19	5	25.7	6.8
35	21	6	21.0	6.0
45	14	14	11.4	11.4
55	15	6	10.5	4.2
65	16	17	10.1	10.8
75	28	9	16.7	10.0
85	28	12	16.1	7.3
95	27	13	15.5	7.5
105	14	8	8.3	4.8
115	10	21	6.3	13.3
125	1	27	0.7	18.9
135	2	25	1.6	20.3
145	2	23	2.0	23.0
155	1	13	1.3	17.6
165	3	3	6.7	6.7
175	0	5	0.0	33.3

By the graphical method parabolic orbits have been derived for the streams (Table III). Table IV gives the distribution of the perihelion distances and the inclinations of the orbits. The intervals in the perihelion distance are one-tenth A.U. except near the earth's distance where they are reduced to 0.05 A.U.; five-degree zones in the inclination are taken near the ecliptic and ten-degree zones elsewhere.

A check upon the accuracy of the results is provided by the work of Guigay (*Journ. des Obs.*, **20**, 132 and 172, 1937) who developed a graphical solution for such orbits, applied it to twenty-eight of these radiants, and checked his graphical results against computations. Guigay's graphical results and the values found here agree equally well

TABLE III

PARABOLIC ORBITS FROM RADIANTS IN *H.C.* 388

No.	λ_1*	β_1	a	A	λ_2	β_2	i	q	Ident.
1	13°	+ 6°	87°	131°	331°	+ 5°	8°	0.43	
2	68	+ 8	37	62	43	+13	156	.29	
3	16	− 5	88	132	332	− 4	5	.47	
4	54	+ 1	50	82	21	+ 2	170	.03	
5	17	− 4	88	132	333	− 3	4	.47	3
6	130	+13	27	46	148	+21	153	.60	
7	328	+ 9	137	165	300	+ 3	3	.95	
8	36	− 9	75	117	354	− 9	18	.22	
9	32	− 8	80	123	349	− 8	16	.28	8
10	12	+34	98	142	325	+22	26	.73	
11	26	−18	87	131	341	−15	23	.47	
12	104	+27	28	47	99	+46	134	.98	
13	52	0	59	96	15	0	0	.01	
14	95	− 8	20	34	83	−15	163	.76	
15	93	−10	27	46	75	−17	157	.59	14
16	132	+ 3	14	24	142	+ 5	174	.85	
17	34	−17	85	129	349	−14	21	.45	
18	46	+ 4	80	123	3	+ 3	7	.29	
19	47	+ 2	82	126	3	+ 2	4	.34	18
20	56	− 4	74	116	15	− 4	9	.18	
21	67	+ 4	64	103	28	+ 5	21	.07	
22	133	− 1	2	3	134	− 1	179	.99+	
23	145	+71	74	116	299	+67	67	.99+	
24	64	+ 9	70	111	23	+10	29	.14	21
25	60	+ 8	75	117	28	+ 8	20	.19	21, 24
26	306	+ 9	168	176	310	+ 3	3	.99+	
27	44	+15	98	142	359	+10	14	.64	
28	145	+29	30	50	151	+49	131	.99	
29	146	− 1	4	7	149	− 2	178	.98	
30	131	+52	53	87	81	+83	93	.98	
31	146	+10	10	17	147	+17	164	.99+	
32	61	+ 5	83	127	17	+ 4	8	.39	
33	84	− 6	60	97	47	− 7	48	.03	
34	138	0	7	12	133	0	180	.95	
35	127	+42	45	75	95	+66	106	.88	
36	139	+23	24	39	134	+37	142	.97	
37	67	− 6	79	123	23	− 5	10	.29	
38	43	−69	96	140	345	−38	39	.96	
39	66	+ 9	83	127	22	+ 8	14	.36	
40	145	+ 8	8	13	144	+12	168	.99+	
41	129	+20	36	60	105	+30	134	.54	
42	109	+26	57	93	68	+32	89	.28	
43	96	+37	72	114	44	+36	57	.50	
44	129	− 8	35	59	105	−13	156	.34	
45	100	− 6	64	104	61	− 7	31	.07	

* Ecliptic coordinates of apparent radiants (λ_1, β_1) and space radiants (λ_2, β_2).

TABLE III—*Continued*

No.	λ_1	β_1	α	A	λ_2	β_2	i	q	Ident.
46	99	+11	71	113	58	+11	26	0.19	
47	153	−18	28	47	136	−30	144	.75	
48	86	+65	91	135	20	+41	44	.90	
49	168	+48	49	81	145	+78	100	.99	
50	90	+ 7	96	140	46	+ 5	7	.49	
51	84	+13	107	149	42	+ 7	9	.75	
52	75	+13	118	156	37	+ 5	6	.88	51
53	223	−11	30	50	243	−17	155	.52	
54	88	−32	108	150	43	−17	19	.83	
55	86	−13	114	154	45	− 6	7	.82	
56	178	−18	28	47	161	−29	146	.72	
58*	120	+ 2	98	142	76	+ 1	2	.62	
59	127	−16	95	139	82	−11	15	.60	
59a	124	+ 6	97	141	80	+ 4	5	.60	58
60	94	+21	145	168	75	+ 7	7	.97	
61	163	+ 9	88	132	119	+ 6	10	.43	
62	141	−32	108	150	97	−17	19	.83	
63	238	+18	24	40	225	+30	147	.85	
64	60	+15	160	174	69	+ 4	4	.99+	
65	241	+17	22	37	229	+25	152	.85	63
66	186	+14	72	114	145	+13	33	.19	
67	173	− 8	85	129	129	− 6	12	.41	
68	190	+ 8	75	118	147	+ 8	17	.22	
69	178	+44	100	143	130	+25	30	.80	
70	223	+ 8	58	94	188	+10	80	.03	
71	169	+47	107	149	125	+26	28	.88	69
72	185	+35	98	142	138	+21	25	.73	
73	223	− 2	60	97	186	− 2	16	.02	
74	235	− 7	51	84	202	−10	124	.05	
75	174	+17	111	152	133	+ 9	10	.80	
76	287	+43	43	71	289	+71	109	1.00	
77	167	+20	116	155	129	+ 9	11	.87	75
78	279	+20	21	35	274	+33	146	.98	
79	212	0	74	116	170	0	0	.19	
80	157	+53	115	155	122	+23	24	.96	
81	166	+27	119	157	129	+13	14	.88	75, 77
82	254	+26	41	69	223	+39	119	.52	
83	326	+52	61	99	36	+64	81	.82	
84	89	+67	111	152	104	+27	27	.99	
85	215	−29	78	121	167	−26	42	.43	
86	183	+29	117	155	145	+14	15	.88	
87	260	+16	46	76	229	+22	123	.21	
88	229	+12	78	121	187	+11	22	.27	
89	338	+ 3	29	49	358	+ 4	174	.45	
90	231	+19	83	125	185	+16	25	.43	

* No radiant was numbered 57.

TABLE III—*Continued*

No.	λ_1	β_1	a	A	λ_2	β_2	i	q	Ident.
91	335	+ 5	23	39	351	+ 9	168	0.64	89
92	331	+ 8	20	34	344	+13	165	.76	89, 91
93	227	+ 1	92	136	183	+ 1	1	.52	
94	359	− 2	38	64	25	− 3	173	.19	
95	231	+11	102	145	197	+ 7	8	.70	
96	270	−12	68	108	229	−13	39	.14	
97	258	+ 7	80	123	215	+ 6	11	.29	
98	329	+16	18	40	324	+27	152	.97	
99	243	+11	97	141	199	+ 7	9	.62	
100	207	+55	115	155	173	+24	25	.96	
101	231	+ 6	110	152	189	+ 3	4	.76	
102	231	+16	113	153	191	+ 8	9	.76	
103	22	+75	76	119	149	+59	60	.98	
104	270	+ 7	80	123	228	+ 6	12	.33	
105	243	+22	120	157	207	+10	11	.88	
106	350	+43	47	78	312	+67	102	.87	
107	259	+23	113	153	219	+12	13	.82	
108	344	+40	48	79	305	+60	102	.76	106
109	281	+67	90	134	214	+42	44	.93	
110	8	+48	48	80	344	+79	100	.99	
111	343	+ 2	51	84	310	+ 4	164	.02	
112	317	+10	79	122	273	+ 9	16	.31	
113	344	− 1	52	86	310	− 1	165	.01	111
114	343	+ 1	54	89	308	+ 1	140	.00+	111, 113
115	308	+19	90	134	263	+14	20	.53	
116	53	+37	38	63	73	+59	117	.92	
117	331	− 6	67	107	291	− 7	24	.09	
118	65	+27	36	60	88	+41	126	.67	
119	35	+36	36	60	33	+60	121	1.00	
120	303	+14	96	140	258	+ 9	12	.62	
121	336	− 7	63	102	298	− 8	39	.05	117
122	349	+12	51	84	316	+16	113	.07	
123	32	+44	45	75	16	+74	106	.99	119
124	46	+34	33	55	52	+56	124	.98	
125	5	−13	38	63	340	−20	143	.31	
126	349	+11	53	87	314	+14	104	.05	122
127	306	+48	94	138	253	+30	35	.80	
128	350	+16	52	86	315	+21	103	.12	126
129	59	+37	40	66	83	+59	114	.88	
130	64	+44	47	78	103	+67	101	.89	129
131	343	+ 3	60	97	306	+ 3	20	.02	
132	345	+ 6	59	96	308	+ 7	48	.02	131
133	61	+38	40	67	85	+61	112	.89	129, 130
134	55	+37	38	63	70	+61	116	.97	129
135	60	+48	49	81	101	+76	97	.96	

TABLE III—*Continued*

No.	λ_1	β_1	a	A	λ_2	β_2	i	q	Ident.
136	55	−14	21	36	65	−26	152	0.91	
137	355	+ 2	50	82	323	+ 3	160	.03	
138	346	+ 6	60	97	310	+ 7	54	.03	131
139	61	+37	38	63	80	+59	116	.92	129
140	272	+16	134	164	243	+ 6	6	.92	
141	305	+68	95	139	246	+37	39	.94	
142	61	+37	37	62	78	+59	117	.94	129
143	342	+ 8	67	107	303	+ 8	29	.09	131
144	12	− 2	52	86	337	− 2	155	.01	
145	349	− 7	76	119	305	− 6	13	.25	
146	69	+14	15	24	72	+22	157	.99	
147	335	+54	90	134	276	+36	40	.83	
148	348	+ 5	79	122	305	+ 4	8	.26	
149	126	−14	61	99	165	−15	55	.08	
150	144	+57	83	127	211	+42	48	.79	
151	339	+ 7	87	131	295	+ 5	8	.45	148
152	344	+ 1	83	127	300	+ 1	2	.36	148, 152
153	352	+ 6	77	120	309	+ 5	10	.26	
154	23	+38	60	97	329	+46	81	.54	
155	340	− 2	89	133	296	− 1	2	.47	
156	354	− 6	76	119	310	− 5	10	.25	
157	66	+37	37	62	57	+61	119	.99	
158	68	+29	30	50	62	+49	131	.99	157
159	68	+26	28	47	62	+48	132	.98	157, 158
160	11	− 5	64	103	332	− 6	27	.06	
161	73	−22	22	37	71	−37	143	1.00	
162	68	−27	28	47	61	−45	133	.97	161
163	77	−20	20	34	78	−34	146	1.00	161, 162
164	28	− 5	69	110	347	− 5	16	.12	
165	26	−17	76	119	341	−15	29	.31	
166	39	+ 5	79	122	356	+ 4	8	.26	
166a	34	−12	85	128	351	− 9	15	.41	17
167	49	− 5	72	114	7	− 5	13	.19	
168	106	+ 2	15	25	96	+ 3	176	.82	
169	54	− 8	69	110	14	− 8	22	.19	167
170	47	−13	77	120	3	−12	23	.29	
171	101	− 7	24	40	85	−12	164	.65	
172	68	+28	60	97	26	+33	78	.31	
173	24	+29	100	143	339	+17	21	.73	
174	307	−17	163	175	306	− 5	5	1.00	
175	44	+15	84	128	359	+12	20	.41	
176	21	+41	102	145	334	+22	26	.84	
177	114	+ 9	16	27	105	+15	163	.88	
178	115	−21	24	40	106	−35	143	.92	
179	48	+29	81	124	0	+24	37	.49	

TABLE III—*Continued*

No.	λ_1	β_1	a	A	λ_2	β_2	i	q	Ident.
180	42	0	85	129	358	0	0	0.39	18
181	108	+ 8	21	35	95	+13	164	.73	177
182	43	+ 5	85	129	359	+ 4	6	.39	18
183	116	+ 9	16	27	106	+15	162	.88	
184	19	+45	106	148	334	+25	28	.85	
185	106	−26	33	58	88	−42	130	.79	
186	38	+46	92	136	345	+31	36	.77	
187	52	+ 2	73	115	15	+ 1	2	.18	
188	43	+21	89	133	357	+16	22	.54	
189	112	−22	29	49	96	−36	138	.80	
190	138	−12	14	24	143	−21	159	.97	
191	143	+12	12	21	144	+21	159	1.00	31
192	55	−17	88	132	11	−14	20	.49	
193	40	+34	102	145	355	+20	24	.76	
194	64	−18	84	128	20	−15	23	.41	37
195	64	−17	86	130	19	−14	21	.47	37, 194
196	335	+30	150	170	334	+10	10	1.00	
197	135	+63	65	105	8	+73	76	.97	
198	113	+64	72	114	15	+61	66	.92	
199	125	− 9	36	60	102	−14	154	.34	
200	96	−11	56	92	54	−14	40	.13	45
201	100	−10	62	101	62	−12	48	.07	45, 200
202	76	+46	89	133	21	+34	41	.73	
203	94	− 8	70	111	54	− 8	23	.14	45, 200
204	146	−34	38	63	125	−54	119	.88	
205	130	+60	66	106	31	+68	74	.92	
206	92	+34	85	129	42	+27	36	.59	
207	84	−32	92	136	36	−23	29	.65	
208	75	−23	104	147	31	−14	16	.73	
209	82	−23	101	144	37	−15	19	.68	208
210	170	− 2	11	19	162	− 3	3	.89	
211	140	−15	47	77	119	−20	123	.19	
212	203	+62	63	101	337	+78	79	.98	
213	200	+58	59	96	325	+83	84	.99	212
214	215	+21	30	50	232	+34	140	.73	
215	183	+19	22	37	174	+31	147	.92	
216	105	+ 8	101	144	62	+ 5	6	.64	
217	187	+ 9	28	47	168	+15	159	.57	
218	260	− 9	45	75	290	−12	142	.13	
219	179	+ 7	39	65	153	+11	156	.25	
220	232	+24	26	44	240	+41	138	.94	
221	141	− 2	96	140	97	− 1	2	.58	
222	158	+ 8	80	123	115	+ 7	14	.31	
223	121	+21	117	155	83	+10	11	.84	
224	149	− 9	90	134	106	− 7	9	.50	

TABLE III—*Continued*

No.	λ_1	β_1	a	A	λ_2	β_2	i	q	Ident
225	143	−11	99	143	99	− 7	9	0.64	224
226	169	+13	74	116	126	+13	29	.23	
227	218	+24	37	62	193	+37	128	.60	
228	154	−11	95	139	110	− 7	9	.59	
229	213	+35	50	82	174	+48	101	.59	
230	163	+31	74	116	134	+30	50	.41	
231	280	+23	36	60	303	+36	132	.62	
232	174	+18	90	134	129	+13	19	.53	
233	153	+50	108	150	108	+24	27	.91	
234	264	+38	39	65	250	+63	116	.98	
235	130	− 8	145	170	106	− 3	3	.97	
236	195	+ 8	85	129	150	+ 7	11	.41	
237	322	− 1	39	65	348	− 2	176	.17	
238	224	0	74	116	182	0	0	.19	
239	270	+38	46	76	235	+56	107	.76	
240	215	+ 5	86	130	172	+ 4	7	.43	238
241	277	+52	56	92	206	+74	88	.92	
242	218	+14	84	128	173	+11	17	.43	238, 240
243	265	+55	64	103	184	+65	77	.86	241
244	255	+53	67	107	183	+58	72	.79	241
245	283	+43	47	78	245	+67	103	.88	
246	205	− 7	100	143	161	− 4	5	.67	
247	223	+11	82	126	179	+ 8	14	.36	
248	223	+17	85	129	178	+13	20	.43	247
249	341	+12	32	54	2	+18	153	.49	89
250	214	+74	96	140	157	+40	40	.96	
251	242	+ 6	82	126	197	+ 5	9	.36	
252	88	+63	104	146	129	+31	32	.93	
253	230	+25	98	142	185	+16	19	.68	95
254	242	+30	90	134	195	+22	28	.60	95, 253
255	316	+23	28	47	303	+38	139	.86	
256	234	+20	98	142	189	+12	15	.68	95, 253
257	323	+15	19	32	314	+26	152	.92	98
258	244	+ 7	89	132	202	+ 5	8	.49	
259	237	+24	99	142	192	+15	18	.69	
260	247	+32	92	136	199	+22	28	.65	
261	315	+17	29	49	296	+28	144	.67	98
262	359	+50	53	87	55	+75	93	.93	
263	237	+24	112	153	196	+12	14	.83	
264	274	−12	84	128	230	−10	15	.41	
265	17	+14	21	36	30	+23	154	.83	
266	346	+20	26	44	333	+33	143	.82	
267	286	+ 2	77	120	243	+ 2	4	.25	
268	34	+53	58	94	102	+70	87	.88	
269	15	+45	46	76	34	+74	106	.98	

TABLE III—*Continued*

No.	λ_1	β_1	a	A	λ_2	β_2	i	q	Ident.
270	276	+38	91	135	226	+25	31	0.68	
271	185	+83	97	141	189	+39	39	1.00	
272	306	+ 7	82	126	262	+ 5	9	.38	
273	311	+ 5	79	122	268	+ 4	8	.28	272
274	304	+13	87	131	259	+10	15	.47	272
275	318	− 1	75	117	276	− 1	2	.20	
276	336	− 2	57	93	300	− 3	45	.01	111
277	348	+22	55	90	308	+29	90	.23	
277a	334	− 1	64	103	295	− 1	3	.05	111, 276

with his more accurate computations (compare Tables III and V). Thus for statistical analysis the orbits indicated in Table III are entirely satisfactory. The most interesting characteristic of Table IV is the marked concentration of orbits with low inclination and perihelion distance between 0.8 and 1.0 astronomical unit.

TABLE IV

Apparent Distribution of Orbits of Meteor Showers

i \ q:	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.83	0.88	0.93	0.98	Sum	
2°	1	3	2	1	1	2	1	1	0	1	0	4	17	39
7	0	1	3	3	4	2	3	3	0	0	1	2	22	
15	1	2	7	5	4	1	4	1	6	2	0	1	34	
25	4	1	3	1	5	2	2	4	2	1	1	3	29	
35	1	2	0	0	1	1	1	1	1	0	2	2	12	
45	1	0	0	0	1	0	0	2	1	0	2	1	8	
55	1	0	0	0	1	1	0	0	0	0	0	0	3	
65	0	0	0	0	0	0	0	0	0	0	1	2	3	
75	0	0	0	1	0	0	0	0	0	0	1	2	4	
85	1	0	2	0	0	1	0	0	1	1	1	0	7	
95	0	0	0	0	0	0	0	0	0	0	1	2	3	
105	0	0	0	0	0	1	0	1	0	3	0	4	9	
115	1	0	0	0	0	1	0	0	0	2	1	2	7	
125	1	1	1	0	0	0	2	0	0	0	0	2	7	
135	0	0	0	0	0	1	1	1	1	1	1	2	8	
145	0	1	0	1	0	0	0	2	1	1	2	3	11	
155	1	0	2	2	0	2	1	0	1	0	1	3	13	
165	2	0	0	0	0	0	1	1	0	1	0	2	7	
172	1	1	0	0	1	0	0	0	0	1	0	0	4	9
177	0	1	0	0	0	0	0	0	1	0	0	3	5	
Sum	16	13	20	14	18	15	16	17	15	14	15	40	213	
									29		55			

TABLE V

PARABOLIC ORBITS COMPUTED BY GUIGAY

No.	i	q	No.	i	q
	°				
1	6.0	0.43	16	174.7	0.79
2	158.6	.33	17	20.6	.43
3	4.8	.45	18	5.2	.33
4	168.	.02	19	3.7	.32
5	4.0	.50	20	8.5	.21
6	151.4	.53	21	16.6	.06
7	3.3	.93	22	180.0	.99+
8	18.5	.20	23	65.	.99
9	12.1	.28	24	27.2	.13
10	24.1	.72			
11	19.0	.44	Perseids	113.6	.96 Comet 1862 III *cf.* No. 139, 142.
12	132.	.99			
13	0.8	.17	Leonids	162.7	.98 Comet 1866 I *cf.* No. 31, 191.
14	166.2	.78			
15	160.4	.61	58	2.	.6
			59a	5.	.61

A STUDY OF TELESCOPIC METEORS

FLETCHER WATSON

Harvard College Observatory

(Introduced by Harlow Shapley)

ABSTRACT

The number of faint meteors that enter the earth's atmosphere each day may be derived from the frequency observed with a telescope of given field diameter, provided the fraction of the atmosphere under effective observation can be determined. This area is related to the path lengths of the meteors, which vary with the meteors' brightness. The probability that a meteor of length L will begin in, end in, cross, or both begin in and end in a field of diameter D is determined in terms of the ratio L/D. From the tables the average path length for meteors of a given brightness can be determined. A scheme is devised by which the magnitude scales for telescopic and naked-eye observations may be joined. From the study of a set of observations the path lengths are found to decrease from twelve kilometers for fifth magnitude meteors to about one kilometer for tenth magnitude objects. Daily frequencies over the whole earth are determined; these show an increase of 2.7 in the frequency for each fainter magnitude. At the ninth magnitude alone the daily rate is approximately one billion (10^9) meteors.

THE study of telescopic meteors yields information on the number, size-distribution, motion, and origin of small particles entering the earth's atmosphere. The number of these particles encountering the earth daily is probably large, but a precise determination of the total daily number within a given range of brightness follows only when the fraction of the atmosphere within the field of observation is known. This fraction depends, however, upon the lengths of the meteor paths, which in turn vary with the magnitude. A method is developed here by which the mean path lengths for various magnitude groups may be found and thence the fraction of the atmospheric surface under observation. The magnitude scale for telescopic meteors is considered and a scheme developed by which the telescopic and naked-eye magnitude scales can be joined without gross errors.

Theory. The theory is based upon the geometry of the effective field—the visible field plus that region in which meteors, moving in any given direction, may begin and have some portion of their path within the visible field. As Öpik pointed out (*H. C.* 355, 1930) the effective field is

$$\pi D^2/4 + DL,$$

where D is the diameter of the field and L the length of the meteor trail. The first term, $\pi D^2/4$, is the visible field of the telescope and the second term, the rectangle DL, is the "invisible field" from which meteors, with part of their path in the visible field, may move. Three

classifications will include all the meteors observed: those that begin in the visible field, end in it, or cross it. A fourth classification, those meteors that both begin and end in the visible field, is a subdivision of the first two classes. In Figs. 1 and 2 the area available for the

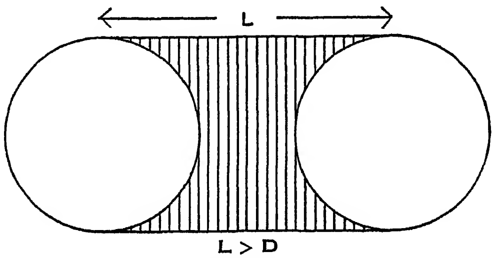


FIG. 1. Effective field for $L > D$.

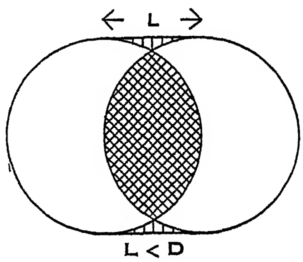


FIG. 2. Effective field for $L < D$.

beginning points of meteors which cross the field is vertically ruled. When the path length is less than the field diameter some of the meteors that begin in the visible field may also end in it; in Fig. 2 the area available for the beginning points of meteors which both begin and end within the visible field is indicated by cross-hatching.

Since the visible field is but a fraction of the effective field, the probability P of a meteor of length L beginning or ending in the visible field is

$$P_B = P_E = \frac{\pi D^2/4}{\pi D^2/4 + DL} = \frac{\pi D}{\pi D + 4L}.$$

The fraction of the effective field available for the beginning points of meteors which cross the visible field, and hence the probability that a meteor of length L will cross the visible field, is given by

$$P_C = \frac{4\left[\frac{LD}{4} - \int_0^{L/2}\left(\frac{D^2}{4} - X^2\right)^{\frac{1}{2}} dX\right]}{D/4(\pi D + 4L)},$$

$$P_C = \frac{4\left[L - \frac{L}{2}\left(1 - \frac{L^2}{D^2}\right)^{\frac{1}{2}} - \frac{D}{2}\sin^{-1}\left(\frac{L}{D}\right)\right]}{\pi D + 4L}.$$

When L is equal to or greater than D, P_C reduces to:

$$P_C = \frac{4L - \pi D}{\pi D + 4L}.$$

The probability that a meteor will both begin and end in the visible field, proportional to the fraction of the effective field crossed-hatched in Fig. 2, is:

$$P_{BE} = \frac{\int_{L/2}^{D/2} \left(\frac{D^2}{4} - X^2\right)^{\frac{1}{2}} dX}{D/4(\pi D + 4L)},$$

$$P_{BE} = \frac{2\left[\frac{\pi D}{2} - L\left(1 - \frac{L^2}{D^2}\right)^{\frac{1}{2}} - D \sin^{-1}\left(\frac{L}{D}\right)\right]}{\pi D + 4L}.$$

P_{BE} vanishes when L exceeds D. From the manner in which the quantities, P, are defined, the meteors which are counted as both beginning and ending in the field are counted also as either beginning or ending in the field, thus:

$$P_B + P_E + P_C - P_{BE} = 1.$$

The probabilities that meteors of length L/D will be observed in these categories have been computed from the equations above and appear in Table I and in Fig. 3. In Table I a column has been added to indi-

TABLE I

CLASSIFICATION DISTRIBUTION OF OBSERVATIONS WITH VARYING L/D

L/D	P_B and P_E	P_C	P_{BE}	Area Effective Field	Area Increase, Doubled D
0.0	1.0	0.0	1.0	1.0	4.0
0.1			0.775		
0.2	0.803	0.003	.609	1.25	3.32
0.3			.451		
0.4	.663	.005	.333	1.51	3.00
0.5			.237		
0.6	.569	.016	.154	1.76	2.78
0.7			.097		
0.8	.497	.062	.056	2.02	2.66
0.9			.024		
1.0	.440	.120	.000	2.27	2.56
1.25	.386	.228		2.59	
1.5	.344	.312		2.90	2.42
1.75	.310	.380		3.22	
2.0	.282	.436		3.54	2.32
2.5	.239	.522		4.18	
3.0	.207	.586		4.81	2.22
3.5	.183	.634		5.45	
4.0	.164	.672		6.10	2.18
5.0	.135	.730		7.35	2.15

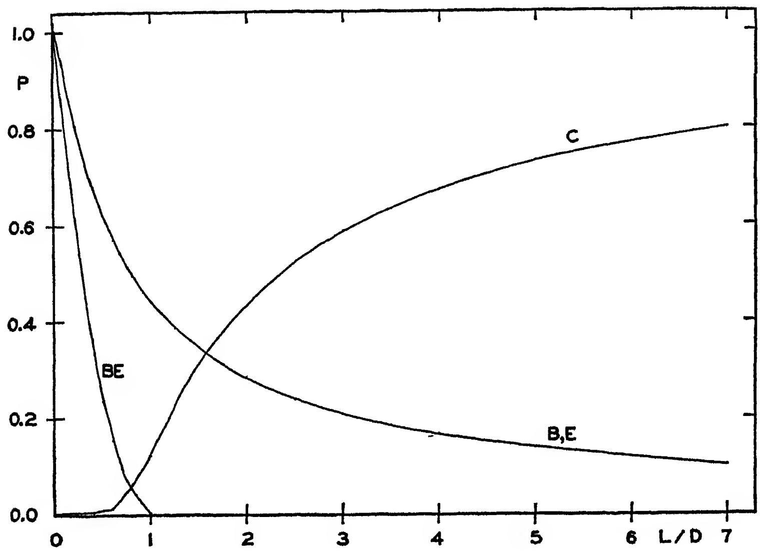


FIG. 3. Probability of classification.

cate the relative effective field for various values of L. In the last column is given the expected increase in frequency for given L/D when the diameter of the observational field is doubled, values of L/D being kept in terms of the smaller diameter. The ratio of frequencies ranges from only 4 to 2 as L/D goes from zero to infinity, hence when this column is used a large number of observations are necessary to yield a reliable value for L/D. From Fig. 3 it is apparent that the length L cannot be determined with much accuracy when it exceeds $4D$. Little weight should be given to a single observation in any classification as it may result from an observational error or from projection effects. Furthermore, sizable numbers of observations are needed to yield reasonably accurate values of L/D.

Study of Observations. In 1934 the writer made telescopic observations of meteors from an elevation of 6000 feet in the mountains of Southern California, on nine nights in July, 8–9 to 16–17, and eight nights in August, 4–5 to 11–12. The four-inch telescope of the A. A. V. S. O., used for the telescopic observations of the Arizona Meteor Expedition, was mounted in a horizontal position with a silvered mirror set at 45° to the horizontal directly in front of the lens. In this manner the zenith could be watched while the observer was in a comfortable position. From declination observations and the duration of star transits the diameter of the field was determined to be 220 minutes of arc. To the observer the eye-field, as seen in the telescope, was 66° in diam-

eter; thus the magnification was 18×. Points in the field were located with respect to a set of cross-wires (three each way), turned at 45° to the horizontal. Their spacing was such that along the diagonals the crossing points were separated by approximately one degree of the sky, or, more precisely, 20° of the observer's field.

The large size of the eye-field suggests that for at least the fainter meteors the observer would not be able to "cover" it. Let us assume that the center of the field was the fixation point for the observer at all times, although certainly for a fraction of the observing time this was not true. To investigate the effective field of view of the observer, the distances of the beginning and ending points from the center of the field were measured for all trails. For various magnitude groupings of the July observations, the frequency of points in rings 10′ wide (3° to the eye) are shown in Table II. For the bright objects, the distribution

TABLE II

VARIATION IN SIZE OF EFFECTIVE FIELD WITH MAGNITUDE

	July									August		
Apparent Magnitude	<3			3.5			4.5			4.5		
Zone	B	E	S	B	E	S	B	E	S	B	E	S
0–10′.	0	0	0	1	0	1	1	3	4	1	0	1
10–20 ..	1	1	2	0	1	1	6	8	14	3	7	10
20–30 ..	0	0	0	4	4	8	6	2	8	6	3	9
30–40 ..	1	0	1	3	4	7	19	17	36	8	3	11
40–50 ..	1	0	1	2	3	5	15	10	25	8	9	17
50–60 ..	0	0	0	5	2	7	4	6	10	5	4	9
60–70 ..	3	0	3	3	3	6	11	7	18	7	6	13
70–80 ..	3	1	4	2	6	8	4	1	5	10	3	13
80–90 ..	4	0	4	3	0	3	3	1	4	4	9	13
90–100 .	0	2	2	3	3	6	2	1	3	5	9	14
100–110 .	0	0	0	0	0	0	1	1	2	8	8	16
Sum.....	13	4	17	26	26	52	72	57	129	65	61	126
Diameter of effective field..........			220′			140′			80′			
Diameter of effective eye-field..........			66°			42°			24°			

indicates that the total eye-field was utilized, although the data are insufficient to be decisive. For objects of apparent magnitude 3.5, we conclude that the effective field is less than the eye-field, and that an effective field of 140′ diameter (42° to the eye) would be approximately

TABLE III

ANALYSES FOR PATH LENGTHS

(a)

$D = 220'$

App. Mag.	$\frac{B+E}{2}$	C	BE	S	$P_{B,E}$	P_C	P_{BE}	$(L/D)_{B,E}$*	$(L/D)_C$	$(L/D)_{BE}$	$\overline{(L/D)}$	L'
1.5	3.5	6	2	11	.31 ±.15	.55 ±.27	—	1.8 0.9– 4.2	2.7 1.7– 7.	—	2.3	530
2.5	5	20	1	29	.17 ±.06	.69 ±.19	—	3.9 2.7– >7.	4.3 2.3– >7.	—	4.1	925
3.5	22.5	12	8	49	.45 ±.10	.25 ±.08	.16 ±.06	1.0 0.7– 1.5	1.3 1.1– 1.6	0.6 0.5– 0.7	1.1	240
4.5	60.5	5	41	85	.71 ±.17	.06 ±.03	.48 ±.10	0.3 0.1– 0.7	0.8 0.7– 0.9	0.3 0.2– 0.4	0.3	66

* Upper figure is value from curve; lower values indicate limits from uncertainty in P.

(b)

$D = 110'$

App. Mag.	$\frac{B+E}{2}$	C	BE	S	$P_{B,E}$	P_C	P_{BE}	$(L/D)_{B,E}$	$(L/D)_C$	$(L/D)_{BE}$	$\overline{(L/D)}$	L'
1.5	1.5	5	1	7							Large	—
2.5	—	22	—	22							>7	—
3.5	13.5	13	4	36	.41 ±.11	.38 ±.12	.11 ±.06	1.1 0.7– 1.9	1.7 1.3– 2.4		1.5	165′
4.5	43.5	2	27	62	.70 ±.11	—	.44 ±.11	0.3 0.1– 0.6	—	0.3 0.2– 0.4	0.3	33′

(c)

$D = 55'$

App. Mag.	$\frac{B+E}{2}$	C	BE	S	$P_{B,E}$	P_C	P_{BE}	$(L/D)_{B,E}$	$(L/D)_C$	$(L/D)_{BE}$	$\overline{(L/D)}$	L'
3.5	3.5	9	1	15	.23 ±.11	.60 ±.25	—	2.7 1.6– 5.6	3.2 1.6– >7		3.0	160′
4.5	25	5	5	25	.50 ±.14	.20 ±.12	.20 ±.12	0.8 0.5– 1.4	1.2 0.9– 1.5	0.5 0.4– 0.7	0.8	44′

correct. The very constant sum over increasing area results in part from occasional displacement of the fixation point from the center of the field. For the faintest meteors the concentration of beginning and ending points to the center of the field is conspicuous, which would indicate that an effective field of 80′ diameter (24° to the eye) is of the correct order of size. In any magnitude group of Table II there is little evidence that the beginning points were systematically closer to the field center than the ending points, or that the beginning points were significantly more numerous in the field than were the ending points. Among the August observations the faint meteors are widely scattered from the center of the field, and we conclude that the fixation point must frequently have been at a considerable distance from the field center. For this reason the remainder of the discussion will be based upon only the July observations.

Let us indicate the number of meteors beginning in the field by B, the number ending in the field by E, the number crossing the field by C, and the number both beginning and ending in the field by BE. For the various groupings in apparent magnitude we may proceed, using Fig. 3, to determine the effective path length L (see Table IIIa). The indicated ranges of solution are found from the natural uncertainties of the counted numbers, *i.e.*,

$$\Delta P_C = \frac{N_C}{N_S}\left[\frac{N_C + N_S}{N_C \cdot N_S}\right]^{\frac{1}{2}}.$$

Since for the faintest grouping nearly one-half of the meteors both appeared and disappeared within the field, a representative minimum path-length may be derived as the mean of the observed lengths. The value of L found in this direct manner is 40 minutes of arc. In the formation of Table III(a) the diameter of the field was taken to be equal for all the magnitude groups, but from Table II we have clear evidence that, at least for the faintest group, the effective diameter is considerably less than that of the eye-field. From the charts used for recording, it is possible to make a new reduction with a field diameter less than that presented by the eye-piece.

Let us assume the new diameter, D_2, as one-half that of the eye-field, 110′, or 33° to the eye. This area is approximately that within the central region of the cross-wires. From the charts the number of meteors observed within this smaller field may be counted. It may be seen from the results in Table III(b) that the reduction in field diameter has not changed the fractional size of L/D for the two fainter groupings, but has halved the path lengths when expressed in arc. On

the observational charts we may once again halve the field diameter, making it now 55′, or 16°.5 to the eye. For such a small field the frequencies have become too small to permit any solution for the brighter magnitudes, but for the two fainter groups the solutions appear in Table III(c). The results are consistent with those from Table III(b) and it is apparent that for the faint meteors the eye-field is 30° to 40° in diameter. The last column in Table I offers a further check upon the general accuracy of the results. In Table IV are given the fre-

TABLE IV

App. Mag.	Frequencies from Table III			N_a/N_b	N_b/N_c	L/D	L/D	L'
	N_a	N_b	N_c					
1.5	11	7	—	1.6 ±0.8	—	—	—	—
2.5	29	22	12	1.3 ±0.4	1.81±0.7	—	—	—
3.5	49	36	15	1.35±0.3	2.40±0.8	—	1.5	80:
4.5	85	62	25	1.37±0.2	2.50±0.6	—	1.2	60:

quencies and their ratios for the various field diameters. From Table I the values of L/D in units of the smaller D are found as indicated. They support the small values of L found in Tables III(a)–(c), although the accuracy is exceedingly low because of the slow change in frequency ratio with changing L/D. In Table IV several values of the frequency ratio fall below 2, which is immediately an indication that the diameter of the effective field is not as large as that of the field involved in the formation of that ratio. Thus even for the second apparent magnitude there is some indication that the entire eye-field is not being covered. The natural uncertainties of the counted numbers prohibit us from taking these ratios as precise, yet they are indicative.

Magnitude System. The combination of the telescopic observations with those made with the unaided eye can be effected only when the apparent magnitudes of the telescopic observations are reduced to the naked-eye scale. During the telescopic survey, simultaneous naked-eye observations of the zenith were made by a second observer. In a number of cases meteors appearing faint to the naked-eye observer were observed through the telescope, thus permitting the magnitude systems to be joined. Since a four-inch telescope theoretically extends the limit of visibility some six magnitudes beyond that of the unaided eye, only telescopic meteors classified as zero or brighter might be expected to be visible to the unaided eye. Such a great difference in the scales is, however, offset by the increase in angular velocity and the

resulting diminution of apparent brightness of the meteors seen telescopically, and also by the failure of the Weber-Fechner law, which is necessarily the basis of the magnitude estimates: very faint objects are classified according to larger changes in intensity than are moderately bright objects. This effect produces a compression of the magnitude scale at the faint end, a compression well illustrated by the magnitudes recorded in the Bonn Durchmusterung. The distribution of the apparent magnitudes of coincident observations is shown in Table V where

TABLE V

Correlation of Telescopic and Naked-Eye Magnitude Estimates

N E Mag. \ Tel. Mag.	−2	−1	−0.5	0.0	0.5	1.0	1.5	2.0	2.5
1.0	1	1							
1.5							?		
2.0								1*	
2.5									
3.0		1				1			
3.5					1	1			
4.0						3	1	1	?
Sum coincidental	1	2	0	0	1	5	1	2	?
Tel. but not N.E.						2	4	7	17
Fraction obs. N.E.	1.0	1.0			1.0	0.71 ±0.60	0.25:	0.22 ±0.14	?
Fraction Öpik	1.0	1.0			0.62	0.50		0.21	
Mean mag. N.E.	1	2			3.5	3.7	4	4	

* This meteor moved very rapidly and had a considerable increase in brightness ten degrees beyond the telescopic field; when crossing the field it was probably near the fourth magnitude, N.E.

interrogation points indicate uncertain identifications. At the bottom of the table are given the mean naked-eye magnitudes of the doubly observed meteors and the fraction seen by the naked-eye observer. For comparison, Öpik's similar ratios are given (*Pub. Tartu Obs.*, XXV, No. 4, 1923). The value of simultaneous observations for standardizing telescopic magnitude scales and luminosity functions is obvious from this reduction.

As a check upon the magnitude scale at the faint end, the apparent magnitudes of various stars were estimated. The differences between their apparent and true magnitudes extended the correction curve to about the eighth magnitude. Beyond that it is an extrapolation to the limiting magnitude. With the telescope employed the faintest star detectable by the writer is 5.8 magnitudes fainter than that seen with

the same eye unaided. The limiting magnitude for the naked-eye binocular observations was 5.8 or fainter, but a difference of 0.3 magnitudes must be allowed between binocular and monocular limits (*Journ. Gen. Physiol.*, 22, 341, 1939). Hence the limiting monocular naked-eye magnitude was 5.5 and the telescopic 11.3. That the faintest observable meteor is brighter than the faintest star by about a half magnitude may be assumed; thus for slow meteors the limiting magnitude was approximately the eleventh. Rapidly moving meteors (recorded as "fast") may have been dimmed another magnitude when observed through the telescope employed (*H. C.* 355, 1930). In Fig. 4 is shown

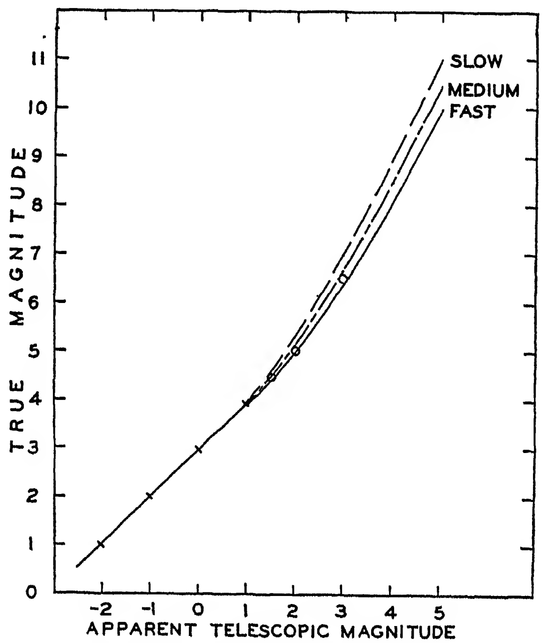


FIG. 4. Adopted magnitude scales.

the relation adopted between the apparent and true magnitudes for meteors of slow, medium, and fast motion. From these curves the mean apparent magnitudes in Tables III(a)–(c) and IV can be transformed into true zenithal magnitudes on the assumption that the mean motion was medium.

The results presented in Tables III(a)–(c) and IV may now be summarized in Table VI. There the mean effective path length in minutes has been given; it is not necessarily the real path length of the meteor, but the path length consistent with the observations and the indicated effective field diameters. Path lengths in kilometers, with-

TABLE VI

SUMMARY OF TRAIL-LENGTH DATA AND FRACTION OF ATMOSPHERE EFFECTIVELY UNDER OBSERVATION

Magnitude		L_{IIIa}	L_{IIIb}	L_{IIIc}	L_{IV}	$L_{\mathrm{Dir.}}$	$\bar{L}$	L_{km}	$D_{\mathrm{eff.}}$	Area Effective	
App.	True									Sq. Min.	Atm
1.5	4.5	500′:	Large				500′:	12	220′	15×10^{4}	18×10^{-8}
2.5	5.8	550′:	Large				500′:	12	220′	15	18
3.5	7.5	(240′)	165′	160′	80′:		160′	5.0	140′	3.8	4.6
4.5	9.5	(66′)	33′	44′	60′:	40′	50′	1.3	80′	0.9	1.1

out correction for projection effects, are computed on the assumption that the height is 86 kilometers. The area of the effective field is given in square minutes of arc and also in parts of the whole earth's atmospheric surface. For the two brighter magnitude groups the figures are in agreement with preliminary results derived from the above and presented in *Harvard Annals*, **105**, No. 32, p. 625, 1937. It should be recalled that these values are for observations made in the zenith whereas the earlier values were corrected to 45° zenith distance. For the two fainter groups the part of the atmosphere under observation is less than in the previous reduction, since here the real effective field diameters have been involved. With these smaller effective fields it is not necessary to include the additional correction of Öpik for the nonrecognition of faint meteors near the boundary of the field. The application of the factors used in *Harvard Annals*, **105**, for this correction gives essentially the same portion of the sky under effective observation as found here; hence the two reductions are in agreement.

The numerical results for L and D are provisional and can best be checked from observations made with less powerful optical equipment. It is expected that such observations, available but as yet unreduced, will soon be investigated.

With the results in Table VI and the magnitude scale in Fig. 4 we are able to determine the number of telescopic meteors of a given magnitude entering the earth's atmosphere each day. Table VII gives

TABLE VII

DISTRIBUTION OF APPARENT MAGNITUDES OF TELESCOPIC METEORS

App. Mag......	−2	−1	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Slow..	0	0	0	1	0	0	3	1	1	0	0	4	2
Medium	1	2	1	1	2	4	5	15	17	20	28	21	2
Fast.....	0	0	0	0	4	1	1	3	6	9	19	11	3

the numbers of meteors recorded in each apparent magnitude and angular velocity interval. Within each group of Table VII a distribution was made according to integral units in the true magnitudes and the frequencies weighted inversely as the effective areas in Table VI. With a small correction for the difference between the total observing time (1529 minutes) and the length of a day (1440 minutes) the data yield the frequencies in Table VIII. Observational incompleteness has appreciably influenced the results for the tenth magnitude.

TABLE VIII

DAILY FREQUENCY OF METEORS OF VARIOUS MAGNITUDES

True Mag.	1	2	3	4	5	6	7	8	9	10
Number	2	2	6	28	53	143	292	790	2240	(1660) $\times 10^6$
Log N	6.3	6.3	6.76	7.47	7.72	8.16	8.47	8.90	9.36	(9.23)

The frequencies increase, on the average, by a factor of 2.7 for each fainter magnitude. Similar results (*Har. Ann.*, **105**, Tercentenary Paper 32, 1937) from a study of the records made by the Arizona Meteor Expedition confirm the order of the frequency of faint telescopic meteors. At the ninth magnitude alone there are approximately one billion (10^9) meteors entering the earth's atmosphere daily.

Summary. From the geometry of the effective field for the observation of telescopic meteors, factors are derived which permit the determination of mean path lengths for meteors of various magnitudes and thence the portions of the atmosphere under observation. From observations made through a four-inch telescope, magnification 18×, it is found that, for the faintest meteors observable, the eye is limited to a field approximately thirty degrees in diameter. A scheme for the reduction of magnitudes of telescopically observed meteors to the naked-eye scale is developed and applied; simultaneous naked-eye observations are important in fixing the bright end of the correction scale. Effective path lengths, consistent with the adopted field diameters, are found to range from over twelve kilometers for meteors of the fifth magnitude to a little more than one kilometer for meteors of the tenth magnitude. Daily frequencies show an increase of 2.7 in the frequency for each fainter magnitude. At the ninth magnitude alone the daily rate reaches approximately one billion meteors.

BINOCULAR OBSERVATIONS OF 718 METEORS

J. D. WILLIAMS

Princeton University Observatory

(Communicated by Henry Norris Russell)

Abstract

The purpose of the paper is to present for study data on faint meteors, secured under circumstances which were optimum to the limit permitted by classical observing methods. The observer was practised, the equipment was well adapted to the problem and sky conditions were excellent throughout.

The principal object of this paper is to make available for study data gathered in a systematic investigation of faint meteors. The decision to give the observational material in so detailed a manner is dictated by the rarity of comparable series and by the limitations of our knowledge in this field. Furthermore, a simple discussion without giving the data would be restricted necessarily to those aspects of the problem which attract the analyzer and his omissions and deficiencies might well reduce the ultimate value of the series.

The writer is acquainted with but two publications which deal with series of this type, due to Öpik and Boothroyd respectively. Öpik [1] gave an instructive discussion of his material but did not supply enough information to enable one to examine certain questions which have since become of interest. Boothroyd's paper [2] is more complete in this respect as some information is given on each meteor, but again lack of certain elements—the observing intervals, for example—restricts its usefulness to others. The present series is numerically larger than either of these.

An attempt is made here to include all the elements of value in the data, but with the following deliberate omissions. The estimates of duration of flight have been suppressed because of the very superior values of Boothroyd which were derived by mechanical means. Color estimates have not been included; they were available for a small percentage of the meteors, but their value is dubious. Finally, it is not possible to deduce radiants in the neighborhood of the field from convergence of paths since the paths have not been sufficiently specified.

[1] *Tartu Publ.*, **27**, 2, 1930.

[2] *H. C. O. Circular* 390, 1934.

The observations were made by Mr. Robert Knabe, a volunteer observer associated with the American Meteor Society. It should be remarked that he was a careful, experienced observer when the program represented by the present observations was undertaken; he had previously plotted approximately 4,000 meteors. If further evidence of his enthusiasm is desired, it may be added that he continued work during at least one earthquake.

Observations from 1934 October 6 to 1935 May 26 inclusive were secured from a station situated at $l = 115°.5$W, $\phi = +32°.8$, while the remainder were made at $l = 111°.0$W, $\phi = +32°.3$. The field of view contained neither clouds nor haze in detected amounts during the observing intervals and observing conditions were not complicated by moonlight or twilight.

The series was secured through the use of a binocular (Zeiss Binoctar) having 50 mm. objectives, magnification 7× and a field diameter of 7°.3, with which stars of magnitude 10.5 could be detected at an altitude of 32°. The definition was excellent over most of the field, but it depreciated noticeably near the edge; the illumination was also somewhat defective there. It would be conservative to confine attention to a smaller field, say 6°, in investigations where these factors are important. Maps of the polar region observed were copied from a photograph and a group of comparison stars was provided. These are listed in Table I where those centrally located and hence most frequently referred to are marked with an asterisk.

TABLE I

Star	Magn.
BD 85° 19	5.0
86° 269	4.5
272	6.0
87° 12	8.0*
88° 2	8.3*
4	6.7*
6	9.7*
8	2.0*
64	7.5
71	6.2

The observer kept his eyes directed toward the pole unless attracted elsewhere by the appearance of a meteor; this circumstance is not trivial since the subjective diameter of the field seen through the binocular is more than 50°. The trail was drawn with a straight edge. Principal emphasis was placed on determining direction of flight and magnitude.

The data are presented in Table II and Table III, the former being heavily coded so as to save space. The following quantities appear there:

S Sidereal time in degrees.

P The angle which meteor's direction of motion makes with the great circle passing from Earth's apex to the center of field, measured counterclockwise. $P = 0°$ indicates flight toward the apex, $P = 270° \sim \odot$.

m Estimated magnitude. Italicized figures show half magnitudes; thus $8 = 8^m.0$, $\mathit{8} = 8^m.5$, etc. Where more than one digit appears variation in luminosity was noticed, the order being chronological.

C In this column the digits have the following significance:

First Digit. Length L of observed portion of path. *E.g.* 4 indicates $3°.5 \leqslant L < 4°.0$, $\mathit{4}$ means $4°.0 \leqslant L < 4°.5$.

Second Digit. Distance r from fixation point to first point of path observed, expressed as above except that $0°.5$ is taken as the unit. *E.g.* 4 means $1°.5 \leqslant r < 2°.0$, 5 means $2°.0 \leqslant r < 2°.5$, etc.

Third Digit. The figures here mean (1) originated in but not contained in field, (2) entered but did not cross field, (3) crossed field, (4) contained in field.

Fourth Digit. (1) casual (*i.e.* not during an observing interval) and implies (2) fixation point unknown, (3) curved, (4) slowed, (5) flickered, (6) diameter increased, (8) diameter large, (9) diameter large but decreased, (0) dummy.

Fifth Digit. Duration of train. Thus $\mathit{0} = 0^s.5$, $1 = 1^s.0$, etc.

Table III lists the observing periods in sidereal time. The date, such as September 3, means on the night of September 3–4, even when work began after midnight. A few intervals during which the writer was the observer are differentiated by an asterisk.

From 114 cases where data are available, the average value of the plotting time per meteor is found to be 50 seconds, standard deviation 10.5 seconds.

An adequate discussion of the material will be undertaken later, but it may not be out of place to exhibit here a few of the distributions which are at hand. The distribution of the magnitudes, based on the final estimates, is appearing shortly in another place.[1]

The angle P promises to be of considerable interest. For instance, since the present observations were made in a restricted field centered at the pole, most of the observed motion was in declination; hence the right ascension of the radiant is quite well fixed unless it is near the pole. Since the right ascension of the radiant is given by

$$\text{R.A. (apex)} - P + 180°$$

[1] *Astronomical Journal.*

TABLE II

S	P	m	C
1934 August 21			
20°	250°	3	6731
23	234	868	47210
28	125	878	3441
1934 September 3			
295°	245°	7	4641
296	254	7	3611
300	251	8	0341
304	315	8	3721
310	274	86	3441
322	240	8	5941
350	189	7	344
351	168	8	2315
358	235	8	1141
1934 September 4			
298°	228°	87	4921
305	35	8	3921
310	272	75	3211
312	357	27	5821
313	256	87	3111
314	293	867	354
323	282	65	274
326	232	5	482
327	201	74	4601
334	216	878	164
336	235	87	344
337	233	7	234
340	163	8	182
350	310	64	464
352	178	7	151
353	337	7	354
356	254	878	334
358	264	87	234
358	275	85	114
0	202	8	144
1934 September 9			
293°	228°	9	134
294	173	76	241
295	45	67	562
296	278	87	163
297	78	87	431
300	255	87	341
302	242	87	441
306	260	76	573
310°	346°	1	56381
312	238	3	183
313	304	867	244
315	232	88	2445
318	274	87	241
321	235	7	284
328	303	65	454
331	220	6	671
332	230	1	68301
332	265	87	284
334	239	88	354
336	311	8	154
338	239	9	482
341	305	87	321
1934 September 10			
298°	259°	68	372
300	346	87	234
304	301	75	42100
305	280	6	264
307	331	87	341
311	242	7	472
311	251	87	1512
313	227	65	251
315	310	6	181
316	290	46	572
319	267	87	161
319	204	979	354
320	250	87	241
322	340	87	234
326	97	98	12400
327	240	87	241
330	280	8	272
330	228	9	254
331	274	7	374
334	231	98	234
335	315	2	48200
336	272	67	282
340	289	87	124
341	351	76	334
342	21	7	124
344	258	76	793
356	29	8	254
10	179	8	683
16	217	7	282
22	98	98	134
25	138	72	421
28	272	867	693
30	288	7	482
37	263	878	274
43	145	8	264

TABLE II—(*Continued*)

S	P	m	C
44°	221°	989	13400
46	106	87	231
47	156	76	251
50	190	7	384
51	193	878	234
51	159	78	214
54	191	4	58201
55	249	6	274
56	249	86	261
57	182	74	311
1934 September 11			
336°	201°	978	214
339	260	878	354
342	290	76	151
343	309	8	224
345	200	3	5828
349	237	54	151
350	233	87	241
1934 September 29			
303°	137°	878	124
304	254	8	472
304	346	65	25400
308	323	4	013
311	237	989	564
313	357	98	144
314	168	76	264
318	307	8	5735
321	270	868	364
322	255	75	241
326	359	778	264
1934 September 30			
306°	325°	989	124
308	299	5	372
314	239	85	241
320	206	76	31100
320	260	67	382
324	254	76	331
327	237	98	434
335	239	75	221
1934 October 6			
338°	249°	62	431
1934 October 7			
312°	344°	8	261
317	249	78	781
323	264	84	331
326	252	878	451
328	331	898	234
331	258	988	254
332	245	85	241
334	166	88	144
336	260	82	4316
341	17	6	551
344	279	7	472
254	244	68	272
0	273	9	261
2	250	5	47300
4	337	85	411
5	291	8	382
7	333	999	334
1934 October 8			
318°	20°	68	372
320	242	878	374
323	248	77	541
332	302	78	472
335	288	9	382
343	247	6	57200
349	245	98	254
352	230	76	421
353	287	8	2448
355	244	7	421
356	335	89	144
1934 October 11			
320°	331°	98	144
327	13	57	272
328	30	8	321
333	312	88	364
338	354	6	361
341	281	57	274
343	24	98	214
345	269	87	331
350	197	7	6838
351	247	867	434
352	305	5	382
352	76	7	173
353	48	8	174
356	218	988	224
358	260	878	682
359	261	98	244

TABLE II—(*Continued*)

s	P	m	C
0°	269°	43	473
1	143	8	1748
1934 October 12			
332°	237°	7	273
339	235	75	16401
342	285	54	48101
344	247	98	441
352	13	87	244
356	276	878	574
8	180	76	341
43	256	88	582
45	58	88	444
46	272	76	561
50	251	988	354
50	276	67	4822
50	231	98	2312
52	280	7	151
56	241	78	582
58	264	4	473
60	192	98	421
63	294	8	374
66	279	65	481
66	229	76	3742
68	267	57	4844
72	195	88	421
72	278	878	354
74	170	65	141
76	200	888	424
78	181	77	25101
81	171	87	231
82	193	7	48201
83	179	8	282
85	270	88	32100
90	191	8	264
1934 October 13			
40°	272°	63	411
40	295	88	331
42	140	2	77382
42	211	8	3822
44	251	87	311
44	271	8	5832
45	197	7	231
52	261	87	241
52	193	98	2412
52	188	57	4722
52	252	8	4612
53	135	988	12400
57	154	87	344
60°	179°	7	382
62	265	76	231
62	255	87	231
67	208	65	57401
72	171	7	4212
73	199	99	214
74	179	87	24101
76	155	6	471
76	257	8	4642
78	159	88	231
85	140	999	32100
91	187	88	344
92	105	9	114
96	183	76	482
1934 October 27			
324°	43°	88	134
325	268	75	461
332	35	44	67380
338	284	88	244
339	255	54	78180
342	214	65	55101
342	286	64	44100
352	227	87	251
358	286	8	251
1934 October 28			
316°	359°	88	154
320	356	989	134
324	304	988	234
326	281	57	373
327	254	88	361
334	290	4	382
337	259	88	3648
338	172	7	15400
342	16	8887	521
344	259	78	38200
356	258	86	411
357	108	988	234
7	184	878	55400
8	248	87	264
9	0	88	224
12	191	8	471
15	23	88	254
1934 October 29			
340°	258°	98	444
341	269	8	264

TABLE II—(*Continued*)

S	P	m	C
346°	242°	7	161
350	263	9	224
350	265	4	683
352	357	87	261
356	269	76	371
358	256	6	482
358	255	989	234
0	278	98	134
2	296	989	234
10	265	88	264
1934 October 30			
342°	7°	88	264
350	59	7	473
356	306	76	351
358	252	87	314
359	298	6	3738
4	209	98	2441
5	181	88	241
6	258	989	384
8	8	99	154
8	265	77	241
10	266	8	482
12	264	5	482
18	195	76	42100
1934 November 1			
338°	267°	88	221
340	304	87	361
342	278	97	144
345	285	989	244
346	269	888	134
347	256	98	154
348	255	878	441
349	30	87	264
352	146	98	321
355	279	78	382
3	315	8	254
4	276	98	164
4	336	8	134
4	340	7	26400
12	159	889	274
12	325	8	214
16	200	7	383
18	276	76	141
19	323	7	382
22	267	87	424
1934 November 2			
56°	203°	7	582
60	262	56	683
1934 November 3			
74°	163°	98	241
76	153	8	47200
80	256	96	221
80	309	98	1542
83	148	8	341
86	217	98	331
86	195	547	58201
90	277	87	271
94	221	87	231
102	226	76	33100
103	184	88	482
106	195	8	28200
108	196	88	382
116	200	87	144
1934 November 4			
326°	352°	8	154
327	234	8	161
327	238	74	31103
336	349	98	154
340	267	75	57300
344	358	989	124
349	296	8	272
354	255	4	78300
358	267	57	582
359	238	87	221
2	295	888	254
6	261	2	783
7	308	87	161
8	354	88	284
9	260	6	781
12	341	988	444
14	205	8	382
20	62	7	161
21	253	5	582
22	258	6	482
1934 November 5			
333°	344°	98	164
342	331	8	182

TABLE II—(*Continued*)

S	P	m	C
346°	11°	87	241
347	23	8	161
351	284	988	144
354	283	98	144
356	203	8	372
358	272	74	221
10	212	999	134
16	346	10 9	114
1934 November 6			
339°	315°	8	272
340	78	4	7638
341	291	98	244
345	179	9	12485
348	348	8	134
350	269	8	282
350	248	9	1442
358	250	74	34100
358	31	888	16460
2	313	8	174
3	269	9	134
5	258	8	144
10	197	8	281
16	260	988	144
17	271	7	674
20	274	98	134
22	237	8	164
28	195	989	344
1934 November 10			
74°	158°	7	58201
75	298	8	183
78	256	8	35100
80	274	7	483
81	191	6	48300
90	219	888	321
92	169	98	374
96	223	98	241
97	197	989	174
99	178	999	154
99	258	8	1512
99	85	8	144
107	128	78	183
107	101	9	234
110	255	84	311
121	186	7	38201
123	168	47	48201
1934 November 11			
4°	283°	77	3821
8	291	88	161
8	358	89	244
12	252	76	431
16	257	6	382
16	267	89	274
22	270	8	144
23	9	8	174
25	303	99	154
32	272	10 9 10	1348
35	305	84	241
35	3	8	2412
41	254	878	22401
41	279	76	1822
48	182	85	43101
52	265	989	144
56	336	87	161
57	315	75	241
63	265	98	344
64	297	8	134
64	258	7	2832
66	191	46	48201
67	263	68	254
68	223	98	384
69	276	989	234
74	184	8	114
76	124	78	68300
78	197	88	151
80	200	8	151
82	195	7	584
88	268	67	464
90	272	7989	674
96	253	8	251
101	226	7	483
113	225	8	151
116	172	54	24102
122	183	56	38203
123	268	8	284
125	202	988	41100
130	87	8	272
1934 November 25			
346°	300°	8	371
356	274	8	5838
0	23	99	114
3	309	78	254
9	271	77	431

TABLE II—(*Continued*)

S	P	m	C
11°	312°	9	134
15	308	8	114
18	264	7	48200
20	278	878	334
21	279	8	592

1934 December 2

S	P	m	C
348°	253°	87	331
353	44	35	58390
354	310	8	241
12	37	78	4725
18	251	989	254
22	299	878	264
37	320	8	124
39	228	7	221
41	284	7	383
47	257	7	371

1934 December 15

S	P	m	C
134°	140°	999	364
134	202	8	151
136	210	10 9 10	324
137	211	8	384
138	165	65	181
144	140	64	32101
146	251	878	114
147	214	8	144
150	136	7	1143
153	143	75	321
155	203	87	151
155	206	8	57220
157	210	9	334
160	137	8	421
165	167	888	324
168	198	878	24402
168	249	8	151
172	59	77	171

1934 December 23

S	P	m	C
6°	61°	7	134

1934 December 25

S	P	m	C
25°	284°	7	483
29	58	8	161
54	252	989	454

1934 December 29

S	P	m	C
89°	329°	88	1611
91	233	989	154
97	253	88	431
104	270	7	2741

1934 December 30

S	P	m	C
14°	340°	87	334
18	349	77	272
96	311	8	251
97	262	989	124
117	311	8	282
120	255	98	241
123	167	7	261
125	287	77	382
138	165	46	282
140	251	65	431
143	293	989	244
144	202	989	282
147	246	8	221
148	113	98	172
148	165	56	1722
150	205	8	382
150	262	67	482
152	211	7	382
154	207	67	382
160	217	8	154

1934 December 31

S	P	m	C
148°	143°	67	482
157	212	888	241
159	150	67	773
159	190	65	25121
168	238	3	282
169	216	10910	154
169	196	8	4822

1935 January 5

S	P	m	C
148°	126°	5	5738
151	121	76	261
154	190	67	382
156	144	86	261
162	246	74	311
167	160	78	382
169	170	7	673
170	132	757	21401
174	312	88	251

TABLE II—(*Continued*)

S	P	m	C
1935 January 25			
63°	238°	88	231
78	301	98	15400
79	35	6	3822
85	290	77	341
88	249	8	182
96	35	7	274
100	125	6	124
107	283	57	482
110	273	878	124
110	272	67	3822
114	182	77	251
114	224	8	261
1935 February 22			
114°	331°	87	371
1935 February 23			
94°	33°	7	241
98	242	76	573
101	268	989	144
107	288	989	144
121	230	68	472
124	112	77	36180
1935 February 27			
102°	247°	6	561
118	218	778	4411
1935 March 12			
213°	254°	78	3711
1935 March 24			
118°	240°	988	134
138	261	87	244
149	254	11	77301
164	293	276	47201
1935 March 25			
133°	262°	8	264
136	232	8	364
137	274	21	78300
140°	286°	61	3618
151	264	88	254
1935 March 26			
120°	36°	89	264
130	247	8	231
141	264	8	231
153	245	76	541
154	271	8	264
162	190	5	37200
163	288	98	241
168	0	8	161
168	262	8	2441
1935 March 27			
134°	277°	989	234
1935 March 29			
185°	268°	88	124
187	264	97	154
192	271	57	382
192	248	86	161
195	244	78	234
199	163	67	67201
202	316	8	1548
203	202	87	33100
206	193	41	36180
213	268	8	372
215	274	22	673
1935 March 30			
146°	262°	77	682
1935 April 2			
140°	358°	7	7838
160	339	98	124
169	247	16	582
170	206	989	214
170	279	5	2732
1935 April 5			
233°	250°	4	2631
237	169	35	5721

TABLE II—(*Continued*)

S	P	m	C
1935 April 24			
159°	250°	76	373
162	227	7	164
164	309	8	251
169	320	56	282
175	217	76	362
184	204	55	1618
188	296	76	451
1935 April 27			
161°	17°	87	251
182	297	8	264
183	271	7	464
193	291	98	244
199	307	7	144
204	298	87	234
207	261	989	334
212	257	8	251
218	266	7	364
233	244	6	272
250	296	6	371
1935 May 4			
166°	281°	57	4721
177	278	78	114
196	294	766	241
198	280	988	154
212	341	6	272
212	199	88	251
215	285	46	472
219	244	75	341
225	310	57	47201
234	248	87	251
246	273	78	272
254	186	7	282
256	201	8	33100
265	218	88	12400
267	283	7	341
272	207	61	34100
275	151	98	251
277	190	8	282
1935 May 23			
195°	302°	52	341
197	246	87	441
198	280	87	331
202	278	7	151
204°	320°	8	244
206	229	7	161
212	292	32	663
1935 May 25			
197°	295°	878	254
206	280	989	254
207	293	5	321
227	260	8	134
230	186	98	144
231	314	878	134
1935 May 26			
232°	189°	76	241
236	311	98	244
239	246	88	334
250	259	7	161
253	241	5	67200
257	285	78	354
264	253	6	251
1935 September 28			
313°	34°	8	134
315	290	6	254
320	300	7	382
327	273	5	372
333	277	88	27400
335	305	88	331
336	301	4	173
345	158	6	171
345	243	87	4412
358	355	7	264
3	183	8	283
9	259	57	482
23	278	7	382
25	162	878	25401
32	197	646	46402
32	191	74	43121
33	205	989	234
41	190	8	282
42	155	87	32100
43	304	7	57201
46	250	76	341
47	200	8	482
48	273	7	144
49	174	74	554
49	265	6	3822
50	158	4	451

TABLE II—(*Continued*)

S	P	m	C
53°	193°	75	*231*
56	224	*8*	*564*
58	263	*7*	*482*
61	170	87	*574*
61	227	6	*2642*
1935 September 29			
293°	277°	85	551
298	5	*8*	124
301	327	*5*	154
305	250	6	673
313	347	6	*151*
313	261	*4*	593
314	256	*7*	*121*
318	311	7	*144*
336	299	*5*	*161*
340	294	*4*	581
345	265	*6*	374
357	192	7	*314*
359	215	7	283
0	247	*7*	*134*
0	255	75	*321*
4	136	8	144
13	247	*7*	*134*
1935 October 1			
33°	247°	87	272
33	221	7	*141*
1935 October 5			
15°	261°	*68*	*482*
15	308	8	*1342*
16	248	75	421
17	265	9*89*	*124*
19	202	*87*	*551*
20	211	7	174
20	193	*87*	*4312*
24	238	*7*	283
24	273	7	*183*
25°	263°	*8*	482
25	273	*76*	*144*
26	349	9	124
32	251	*8*	*221*
33	294	6	273
34	248	*8*	*372*
35	263	3	473
35	186	*7*	*582*
37	180	*7*	*261*
40	177	8	*311*
54	190	34	28202
55	305	*8*	214
56	161	*88*	*324*
59	217	*6*	*25100*
60	218	*7*	561
62	193	8	461
63	198	*8*	*683*
63	266	7	*382*
64	215	*7*	*124*
64	193	*7*	*134*
65	185	7	*26101*
69	349	8	*244*
70	248	*67*	*382*
79	227	*7*	484
1935 October 13			
299°	260°	2	*3831*
1935 October 26			
23°	277°	*8*	321
25	170	*75*	*23100*
29	176	87	224
31	186	5	*58200*
33	285	888	*134*
35	201	26	*78301*
35	338	7	2412
39	329	7	*274*
87	245	*7*	*5831*
89	206	67	482
95	244	8	*683*
96	257	*7*	371
101	205	5*2*	431

TABLE III

Date	Began	Ended	Date	Began	Ended
1934			December 30...	12°	21°
September 3...	343°	356°		92	130
4...	314	340		136	163
	347	5	31...	148	172
9...	290	321			
	326	343	1935		
10...	290	344	January 5...	148	174
	352	356	25...	55	91
	9	59		95	120
11...	336	352	February 22 ..	77	114
29. .	298	335	23.	81	112
30...	304	335		117	124
October 6...	335	340	27...	84	112
7...	310	344	March 24...	113	142
	349	7		146	164
8...	315	0	25...	113	145
11...	318	2		150	165
12...	330	8	26..	115	146
	40	96		150	170
13 ..	38	80	27...	116	134
	83	97	29...	183	193
27...	319	326		198	218
	336	2	30...	132	155
28...	315	348	April 2..	124	155
	351	18		158	176
29..	336	20	24...	152	178
30..	335	1		180	191
	4	19	27...	158	193
November 1...	337	6		197	233
	9	23		238	253
2 .	48	62	May 4...	166	201
3..	72	118		209	247
4...	323	354		253	282
	357	22	23 ..	187	221
5 ..	333	6	25...	196	218
	8	22		220	234
6...	336	10	26...	220	259
	14	28		262	272
10. .	73	83	September 28...	311	348
	89	114		356	12
	118	126		21	64
11...	5	52	29...	290	305*
	55	101		307	312*
	109	132		334	335*
25 ..	344	26		336	8*
December 2..	348	22		13	20*
	29	48	October 1 ..	31	37*
15...	130	172	5..	11	45
23 ..	2	32		54	80
25...	7	40	26...	22	41
	43	63		87	104
29...	90	101			

approximately, the distribution of P during the night gives a good representation (except for a constant) of the distribution in right ascension of the radiants represented. Further, something may be learned concerning richness of the streams from the clustering tendency of the values of P. Mere inspection of the distributions from night to night convinces one that a considerable percentage of the objects listed are not completely isolated, though it is necessary to devise a suitable test of significance before the question can be examined in detail.

TABLE IV

	10°	30°	50°	70°	90°	110°	130°	150°	170°	
P	13	18	8	5	4	6	13	24	36	
A	2	6	7	19	47	76	113	102	108	
	190°	210°	230°	250°	270°	290°	310°	330°	350°	Total
P	64	57	54	121	129	64	49	23	30	718
A	77	55	51	33	7	5	5	1	4	718

The distribution of P for the entire series is given in the first line of Table IV. This distribution is related to that of the space directions of the meteors, but it is not immediately applicable to the determination of the latter since its form depends on the distribution of observing time. It is also subject to strong perturbations other than the obvious ones depending on Earth's motion, two of which have been mentioned in another place.[1] These perspective effects are most naturally referred to the angle of flight A, measured from the vertical circle, which may be found from

$$A = P + \text{Sid.T.} - \text{R.A. (apex)}.$$

The distribution of A is given in the second line of Table IV. It will be noticed that the general course of variation is accounted for, to a large extent, by these effects,[2] though the theory was over-simplified to warrant close comparison with observation. The correlated angle P exhibits a maximum near $P = 270° \sim \odot$ as expected, though its strength seems surprising. Öpik's distribution of P similarly shows a maximum toward the sun.[3] It will be necessary to investigate the theory of directions extensively in order to establish how much real excess there is, if any. Boothroyd's angle A is distributed in such a

[1] Cf. following paper.

[2] *Loc. cit.* Table I, last line.

[3] *Tartu Publ.*, **27**, 2, p. 7, 1930.

remarkable fashion that it is not clear how one may safely use the related distribution P in a general discussion of directions.

In Table V the meteors seen during observing intervals and the intervals themselves are scattered against zenith distance of apex and right ascension of apex, which serves to show the extent of the series in time. (Here 165° stands for $160° \leqslant z < 170°$, etc.; 75° stands for

TABLE V

R.A. (apex) \ z	165°	155°	145°	135°	125°	115°	105°	95°	85°	75°	65°	55°	45°	35°	25°	15°	Total
75°						8	14	14	24	6	7	2	4	3	9		91
						20	31	35	49	24	17	13	13	13	12	2	229
105°					1	34	31	22	10	11	19	19	24	26	9	1	207
					7	115	113	77	31	29	32	52	40	47	36	7	586
135°					34	41	40	14	9	8	11	10	14	9	11	1	202
				2	100	141	114	50	31	36	36	32	22	25	34	3	626
165°				7	9	2	2						7	11			38
				41	31	13	9						16	26			136
195°		1	4		1			4		3	1	10	11	7			42
		2	51	22	13	4		16	13	13	8	19	34	23			218
225°			1	2	2	2	5										12
			14	13	10	12	12	2									63
255°		1	4	3													8
		44	38	16	5												103
285°		2	8	5	7	3	2	6	3								36
	1	65	63	44	50	12	8	8	14								265
315°			6	3	5	5	5	2	1	3	5						35
		4	38	32	32	14	26	25	18	18	16	2					225
345°				5	5	1	5	2	2								20
			2	23	24	24	18	12	9	7							119
Total		4	23	25	64	96	104	64	49	31	43	41	60	56	29	2	691
	1	115	206	193	272	355	331	225	165	127	109	118	125	134	82	12	2570

$60° \leqslant$ R.A. (apex) $< 90°$, etc.) In each of the pairs of rows the number of meteors occupies the first line and the number of degrees in the corresponding observing intervals is in the second. The data are not well distributed for examination of seasonal change in frequency, but a fair set of values can be obtained as a function of z if we neglect the seasonal effect. Hourly frequencies f based on the last two rows are

TABLE VI

z...........	150°	130°	110°	90°	70°	50°	30°
f...........	1.3	3.0	4.7	3.3	5.0	6.8	6.4

listed in Table VI; these were computed making due allowance for the plotting time. The general increase in frequency with decreasing z is strongly evident, but when the points are plotted one is not encouraged to fit to them anything more complicated than a straight line. If we compare the slope of the best line with those of Hoffmeister's variation curves [1] at $z = 90°$, we find agreement for the case where the speeds of the meteors are in the neighborhood of 3, or 3.5 (in terms of Earth's speed), though this comparison seems rather forced since the observed variation curve does not resemble the computed one. Perhaps the comparison would be improved if the theory were modified for observations in a small region at fixed zenith distance.

It is a pleasure to acknowledge my indebtedness to Robert Knabe who volunteered to make the observations and to the Steward Observatory of the University of Arizona, the facilities of which were constantly at our disposal.

[1] *Veröff., Berlin-Babelsberg*, 9, 1, p. 21, 1931.

SYSTEMATIC BIASES OF METEOR DIRECTIONS

J. D. WILLIAMS

Princeton University Observatory

(Communicated by Henry Norris Russell)

(*Read April 22, 1939*)

ABSTRACT

Data concerning the apparent directions of flight of meteors have been utilized principally in the determination of radiants, but the possibility of deducing from them some of the characteristics of the spatial- and velocity-distributions is now receiving attention. This note serves to draw attention to two circumstances which affect the distribution of apparent directions in a systematic manner; one effect occurs only when observations are made in a restricted region of the sky and consequently is of interest when discussing telescopic meteors. The biases are illustrated quantitatively by consideration of an idealized distribution. Both effects enhance the number of meteors whose altitudes decrease during flight.

WHILE the apparent directions of flight of meteors have been used in the past principally in radiant determination, currently there is a tendency toward utilizing these data in more general statistical inquiries, notably in a new indirect attack on the problems of spatial and velocity distribution. So far no general procedure has been published, nor will it be attempted here, but two sources of systematic bias will be adduced which will together produce a first order effect on the distribution of this characteristic. These biases apply primarily to observations made in a restricted field.

We shall assume that all space directions of meteors relative to the point of observation occur with equal frequency. The assumption is not made because this type of uniformity possesses any inherent likelihood, but only because it permits us to exhibit the effects in a simple manner.

The first, which we shall designate by *horizon attraction*, is a simple though very strong effect due to perspective which may be evaluated in the following manner. Suppose that observations are made at zenith distance ζ and at a point O in the atmosphere. Consider a unit hemispherical shell centered at O with its equator contained in the horizontal plane through O. Then the distribution assumed above may be characterized by allowing meteors to fall normally from the surface toward O, the probability that a meteor will appear being the same for all elements of surface. By integration over suitable regions on the

surface we may find the probability p associated with the apparent direction of flight A (or a), where the latter is measured from the vertical circle through O and in a plane perpendicular to the line of sight, so that $A = 0°$ indicates flight toward the zenith. It is not difficult to establish that

$$p(A, \Delta, \zeta) = \frac{1}{2\pi} \int_{A-\frac{\Delta}{2}}^{A+\frac{\Delta}{2}} \int_{0}^{\theta'} \sin \theta \, d\theta da, \tag{1}$$

where

$$\cos a \tan \zeta \tan \theta' = 1$$

and Δ is the class-interval into which the data are grouped. The integrations are easily performed and, for a particular Δ and ζ, lead to the probabilities listed in the first line of Table I.

TABLE I

$\Delta = 20°$. $\zeta = 57°.5$

A	10° 350°	30° 330°	50° 310°	70° 290°	90° 270°	110° 250°	130° 230°	150° 210°	170° 190°
$p \times 10^4$	90	109	164	299	556	813	947	1002	1021
$\overline{L}$	1.00	1.11	1.32	1.72	2.13	2.28	2.27	2.24	2.23
f	100	104	113	128	144	150	150	148	148
F	10	13	21	43	89	136	158	165	168

From this we can appreciate the order of magnitude of the effect: the directions grouped about 170° or 190° have probabilities more than 11 times greater than those near 10° or 350°. As a consequence, the frequencies of observed directions will show a fictitious maximum of great-intensity toward the nearest point of the horizon, which explains the choice of name. It may be remarked that the effect is in no way maximized by the assumption of uniform distribution, so the computed probabilities are not extreme values.

The second, or *length-effect*, is based on the circumstance that the frequency of directions of meteors of given true length of path depends on the apparent length L. In order to develop the idea, consider meteors of given length, say unity, passing normally through the elements of the surface previously described. The average apparent length for each of the categories of A (again for the case of uniform distribution) is

$$\overline{L} = \frac{1}{2\pi p} \int \int \sin^2 \theta \, d\theta da, \tag{2}$$

where the limits are as in (1). Direct integration is not simple here but the approximate values in the second row of Table I, suitably normalized, are readily obtained.

Having established that the apparent length depends on direction, we are in a position to show that the frequencies will be affected. Since, when observing in a restricted field, we may see meteors which do not originate in the field, it is obvious that the effective size of the field is larger than it would be if point objects were under observation. Öpik[1] has expressed this quantitatively in the relation

$$\text{Effective area} = \pi D^2/4 + DL, \tag{3}$$

where D is the field's diameter. La Paz[2] found this to be a good approximation (by replacing $\sin D/2$ by $D/2$ in the second term) for phenomena on a spherical surface. Since in telescopic observation of meteors the field is small it suffices to consider the problem in a plane and in this case equation (3) is exact when the randomization is carried out in this sense: choose at random an element of area and from it as initial point draw a length L in an arbitrary direction. A remarkable feature of the formula is that it is also exact if the meteors all travel in the same direction. This last fact permits us to complete the argument regarding the second bias for we may apply the formula independently to each of the categories in Table I, finding in this manner different effective areas and hence different frequencies for each category. As an example, if the length of meteors in the class centered at $A = 10°$ is half the field diameter, then the frequencies are proportional to the relative numbers f in the third line of Table I. In the last line horizon attraction and length-effect are combined to give the relative numbers F.

If we take the field of observation at the north celestial pole the following expectations are reasonable: for a series of observations throughout the night and year the average position of Earth's apex will be toward $A = 270°$, and on the assumption of uniformity as defined above but allowing for Earth's motion we would have a concentration of directions toward the antapex, *i.e.* toward $A = 90°$. The effects mentioned here will cause too many to be observed in the range $90° < A < 270°$ with a maximum at $A = 180°$; because of the strong biases and the reduced apex-effect accompanying hyperbolic velocities the observed maximum may be very near to the latter one. If we had discussed the angle of flight P relative to the apex direction, the pre-

[1] *Tartu Publ.*, **27**, 2, p. 6, 1930.

[2] *A. N.*, **246**, p. 201, 1932.

ferred directions would have been 180° if undisturbed and 270° = $\odot$ if strongly biased, though the effects would be smoothed out to some extent in this reference system.

Of course one should not expect too much numerically from results based on the assumption of uniformity, but for observations toward the pole it illustrates quite well the type of variation to be expected unless, for instance, the surface density (in our model) is greater at the pole than elsewhere, which is a distribution in no way suggested by theory or observation.

ELECTRONS IN METALS

W. V. HOUSTON

California Institute of Technology

(Communicated by Robert A. Millikan)

ABSTRACT

The study of the theory of metals is part of the program for understanding the properties of material bodies in terms of the behavior of the constituent electrons and atomic nuclei. During the past decade this part of the program has met with considerable success in its development along two lines.

The application of the Pauli exclusion principle and the methods of quantum statistical mechanics has led to an understanding of the absence of a large electronic specific heat. It has made clear how there can coexist the properties ordinarily associated with free electrons and those associated with electrons bound in atoms.

The application of quantum mechanical methods to the study of the motion of electrons in the periodic field of the nuclear forces has brought out the explanation of other combinations of these apparently exclusive properties. In particular, the existence of a current at vanishingly small field strengths indicates that at least some of the electrons are essentially free to move. On the other hand, the ponderomotive force on a conductor carrying a current in a magnetic field, together with the smallness of the Hall effect, indicates a very intimate binding of the electrons to the nuclei. This freedom to move under an electric field, combined with the rapid transfer of momentum to the nuclear lattice, is exhibited by the quantum mechanical behavior of an electron in a periodic field. The states in which an electron has a definite energy can represent a definite average current, and yet the electron momentum will be indeterminate and intimately shared with the lattice. This fact has tended to obscure the position of the conservation of momentum in this theory.

MUCH of present day physics is a study of the properties of matter. In this study there are two types of work that must be carried on more or less simultaneously. The first is the subdivision, or analysis, of matter into smaller and smaller constituents. This is the familiar process of taking it apart. The possibility of building a satisfactory theory of matter, in this way, was recognized by the Greeks, but it was first effectively carried out during the nineteenth century, in the identification and description of the chemical atoms. During the first third of the present century the analysis was carried further. The fundamental particles investigated were the electrons and the atomic nuclei. At the present time there is tremendous activity in pushing this analysis even to the component parts of the atomic nuclei.

Parallel with this work of analysis, must go the corresponding synthesis, the attempt to explain the properties of macroscopic material bodies in terms of the properties and the interactions of their elementary parts. At the stage when the chemical atoms were regarded as the

fundamental particles, the kinetic theory of gases was the most successful attempt at synthesis. Although efforts to understand the behavior of solids and liquids were not very fruitful, many of the properties of gases could be well understood in terms of atoms, acting on each other with short distance forces. Later, with the isolation of electrons and atomic nuclei, the problem of constructing atoms and molecules out of these smaller constituents, was vigorously attacked. The exigencies of this work played a more important role in the development of the quantum mechanics, than did the study of the individual particles themselves. With the development of quantum mechanical methods, the properties of atoms and small molecules may be said essentially to have been mastered, but there still remains much to be done before the same thing can be said about solid and liquid bodies of ordinary size.

The fundamental justification for the conclusions reached in analyzing matter into its constituents, is to be found in the successful, theoretical, construction of the original ordinary bodies out of those constituents. If a suitable, detailed and quantitatively adequate, picture of an ordinary piece of matter cannot be obtained in terms of the elementary particles and their interaction, the utility and even the correctness of the concept of the elementary particles may well be seriously questioned.

It is now believed, rather generally, that such a construction is possible, and that the difficulties are merely those of complexity. However the problem is not one to be ignored on that account. Whether tremendous complexity can introduce qualitatively new elements, is a question of much importance, particularly in biology. It has been frequently discussed, but never very satisfactorily answered. Probably the only satisfying answer will be in an adequate theory of the complex bodies.

During the last decade, the theory of solid bodies has made tremendous strides, and I should like to emphasize briefly, a few aspects of its development, with especial regard to the properties of electrons in metals.

The electron theory of metals obviously dates from the discovery of electrons. Much of the pioneer work was done by Professor H. A. Lorentz and his contemporaries. However it was never possible to construct a system of charged particles, obeying the classical mechanics and electrodynamics, that could suitably represent a metal. The best that could be attained was the qualitative picture of a metal containing a number of electrons free to move under the influence of an external field. This property distinguished the metals from other substances, in which all of the electrons were bound to atoms. Even this simple pic-

ture was beset by difficulties, and it was not until the development of quantum mechanics, during the years after 1925, that it became possible to make an effective attack on the problem.

It is convenient to regard the contributions of quantum mechanics, in this field, under two headings. In the first group may be placed the results of the use of quantum statistics, based on the Pauli exclusion principle. In the second group, may be placed the results of the quantum mechanical treatment of the motion of electrons in the nuclear fields.

The Pauli exclusion principle was first applied to this problem by Pauli himself, and then by Sommerfeld. As a result of Sommerfeld's work, there was cleared up the long standing problem of the apparent absence of any specific heat attributable to electronic motions. The classical statistical mechanics indicated that each independent particle in a complex body should contribute to the specific heat just three times the Boltzmann constant. This was in fair agreement with the observations if one counted only the whole atoms. But if the atomic nuclei, and all of the electrons, were counted separately, the specific heat to be expected became many times as large as that observed. Even if there were counted only enough free electrons to explain the observed electrical conductivity, the result was still quite clearly too large. These facts seemed to indicate that although it was necessary to postulate free electrons to explain conductivity; to understand the specific heat, it was necessary to regard all the electrons as so tightly bound to the atoms that they had no independent motion of their own. Although the electrons could be moved easily by an external electric field, they apparently could not acquire energy by collision with other electrons or nuclei. Such a situation was a definite contradiction of the principles of classical mechanics, and was for many years a serious obstacle to the development of a theory of metals. This obstacle was finally surmounted by the application of the Fermi-Dirac statistics, based on the Pauli exclusion principle.

The Pauli exclusion principle is involved in all applications of quantum mechanics to complex systems, but it appears, in a sense, to be extraneous to the mechanics. The principle was first formulated as a result of the study of complex spectra, and the attempt to explain them in terms of atomic structure. The original statement was that no two electrons can occupy the same state in an atom. This simple statement must be interpreted in terms of the approximation methods used for treating a complex atom and its spectrum. The precise statement is that the Schroedinger wave function, which describes the system, must

be such as to change sign if any two electrons are interchanged. It seems worthy of remark that this principle is not a property of an isolated electron, but is a property of pairs of electrons. It could never have been discovered by studying one electron by itself, but only by studying their behavior in groups. The possibility that there may exist other principles that become effective only when still larger groups of electrons are considered, adds to the importance of the theoretical study of complex systems.

The first suggestion that the exclusion principle could play an important part, and must be applied, in the theory of metals, was met with considerable incredulity. It was felt to be reasonable that electrons in the same atom should be closely associated with each other, and that they might well exclude each other from the various atomic states; but it seemed fantastic to say that electrons in different parts of a large piece of metal must be required to be in different states. The difficulty was due largely to the use of the original formulation of the principle, which was devised for spectroscopic theory. When the principle is stated by saying that no two electrons, with parallel spins, can be at the same place at the same time, it becomes evident both that the principle must be applied to all of the electrons under consideration, and that it really imposes no restriction on electrons that are far apart. If it could be stated that no two electrons at all can occupy the same position at the same time, the restriction would seem to be merely a part of the concept of an individual particle. Even as the statement stands, the necessity of applying it to all the electrons in a complicated system is obvious.

When the restriction imposed by the Pauli exclusion principle is combined with the quantum mechanical principle of indetermination, it follows that to pack a group of electrons close together, requires that they be given energy. This has nothing to do with their potential energy of mutual repulsion, but is in addition to it. The necessity for this kinetic energy exists, even for electrons in a metal, where their electrostatic fields are largely neutralized by the positive ions. The requirement that all electrons have different positions, combined with the fact that a number of them are packed into a small space, involves a considerable degree of determinateness in position. This is accompanied by a corresponding indeterminateness in momentum and kinetic energy, and, therefore, a large average kinetic energy. When electrons are packed as close together as they are in a metal, the average kinetic energy that must be given them, in accordance with the Pauli exclusion principle, is so much greater than the average temperature energy to be expected, that the influence of the temperature on the motion is negli-

gible. The electrons are free to move under an external field, but they are also bound, in the sense that their mean kinetic energy of agitation is determined by considerations that are quite independent of the temperature. If an electron collides with a nucleus, the energy that it can gain is negligible, compared with its energy due to the requirements of the exclusion principle; and the amount it can lose is similarly small. The calculation shows that the specific heat of a gas of free electrons, packed as closely as in a normal metal, is proportional to the temperature. At ordinary temperatures it amounts to about 2% of that to be expected on the classical theory. This fact, which was developed by Sommerfeld, eliminated the old difficulty, and brought about a revival of interest in the theory of metals.

The other class of applications of quantum mechanics includes detailed studies of the motion of electrons in the field of the nuclei. According to the classical theory, the electrons move among the atomic nuclei, exchanging energy and momentum with them in collisions. An external electric field causes the electrons to move to the boundaries of the metal, and to accumulate at the boundaries in sufficient numbers to reduce to zero the effective field inside. At the surface of the metal the electrons are retained by a potential barrier. The origin of this barrier is not very clearly understood, but it is presumed to be associated with the attractive forces due to the positive ions. This crude description of the interaction between electrons and nuclei, simple collisions inside the metal and a barrier at the surface, is adequate for some purposes. It gives a fairly satisfactory account of the emission of electrons from hot bodies, their emission from cold bodies under the influence of strong external fields, and the ordinary surface photoelectric effect. With the additional assumption of a mean free path of the electrons between collisions, it is possible to develop a theory of conductivity and the thermo-electric phenomena. However, a complete theory must be able to give a detailed account of the whole phenomenon, and must be able to provide the basis for a quantitative determination of the mean free path. Such a program requires a more detailed study of the interactions.

In addition to its inadequacy in details, the classical picture meets a difficulty in accounting for the ordinary ponderomotive force on a conductor carrying a current in a magnetic field. Consider in particular the arrangement shown in Fig. 1. A long strip of conductor, S, is connected by two sliding contacts, C_1 and C_2, to a battery. The ends of the strip are so far away that they can be neglected. In the strip the current will flow along some such paths as indicated. Now let there

be established a magnetic field, perpendicular to the plane of the figure. The most prominent effect observed is a force parallel to the length of the strip which must be explicable in terms of forces on the electrons.

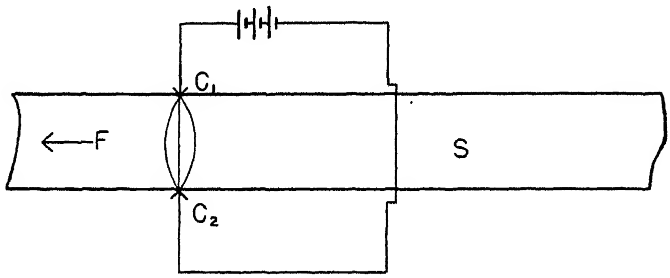


FIG. 1. Ponderomotive force on a conducting strip. S is an infinite strip of conductor free to move parallel to its length. The current flows through the conductor between the points C_1 and C_2 along paths such as are indicated by curved lines. In the presence of a magnetic field perpendicular to the plane of the figure the strip experiences a force F.

The electrons that are moving in the strip will experience a force perpendicular to their direction of motion. As a consequence, one expects an additional current to be set up along the length of the strip. The momentum of this current would then be transferred to the metal through the mechanism of resistance, and the force on the strip could be understood. This would be the normally expected explanation of the force. The trouble with this explanation is that the current set up along the strip is not usually of the amount to be expected. Its magnitude is given by the coefficient of the Hall effect, and this coefficient is sometimes negative, as is to be expected for negatively charged electrons, but it is sometimes positive. The current set up by the action of the magnetic field is sometimes in the direction to account for the observed force, but it is sometimes in the opposite direction. Nevertheless the force is always the same. It is quite independent of any specific property of the metal such as the Hall effect, or the resistance. One can say that the force seems to be exerted directly on the conductor, rather than on the current flowing in it.

The interpretation of these facts requires that there exist a very intimate connection between the free conduction electrons, and the positive ions of which they are supposed to be free. There is evidently a very rapid transfer of momentum between the electrons and the atomic nuclei, and the transfer is essentially independent of that involved in the resistance.

The quantum mechanical treatment of electrons in a periodic potential field provides a satisfactory basis for understanding such a

situation. Consider, for example, an electron moving in a one-dimensional periodic field such as is shown in Fig. 2. This may be regarded as an idealization of the field of the regularly arranged ions in a crystal.

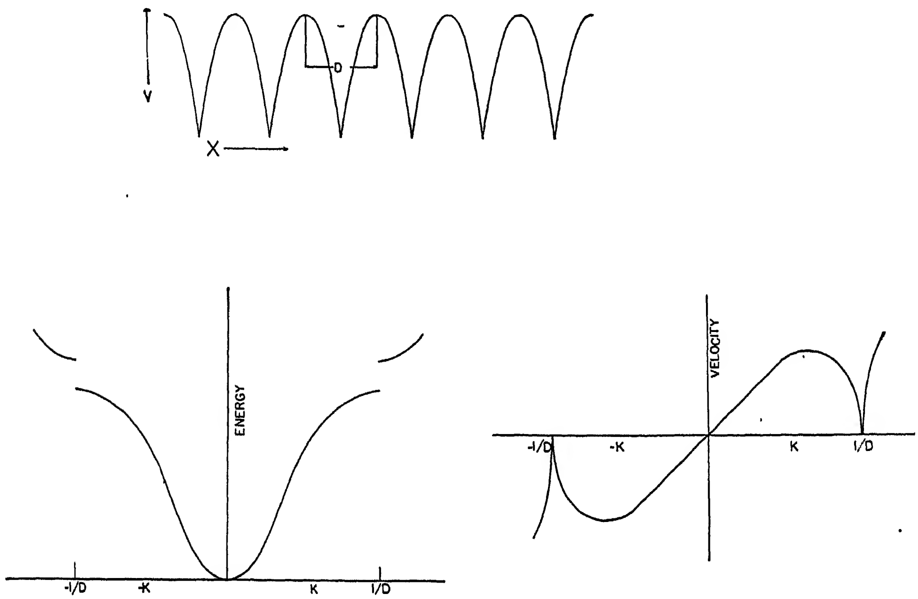


FIG. 2. Electron in a periodic potential field. In a one-dimensional potential field such as indicated at the top of the figure, the states of an electron are designated by a number K. The energy as a function of K is shown at the left and the average momentum or velocity is shown at the right.

The states in which an electron has a precisely defined energy can be designated by the values of an index K, and the dependence of the energy on K is shown. For values of K in the neighborhood of zero, the energy is proportional to K^2, but as K increases the energy increases more slowly to the value $K = 1/D$, where it takes a sudden jump. Energy values that lie between these two branches of the curve are not permitted at all. Although the energy of an electron in one of these states has a precisely defined value, the momentum does not. If the momentum of the electron is measured, it cannot be predicted in advance what value will be obtained, but if a measurement is made on a great many different particles in the same state, the theory can predict the distribution of the results. In particular it can predict the average of the results that would be obtained. There will be a definite average momentum associated with each value of K. The figure shows a typical form of the dependence of this average momentum, or the current which is proportional to it, as a function of K. The curious thing is that the

average momentum does not continue to increase, as does the energy. For small values of K, where the energy is proportional to K^2, the velocity is proportional to K. As the energy begins to increase more slowly, the momentum increases more slowly and finally decreases to zero for $K = 1/D$. It is in this relationship between the energy and the velocity, that is found the intimate exchange of momentum between the electrons and the positive ions.

In the presence of an external force, the state of each electron changes. This change can be described as a change in K, since the state is completely described by K. The detailed analysis shows that the quantity K changes just as though it were the momentum of the electron. The rate of change of K is just equal to the force divided by Planck's constant h. But for many values of K, an increase in K actually represents a decrease in the electronic momentum. Since momentum is of course conserved, this can only mean that momentum is being given to the lattice of positive ions. Even though the force is such as to act on the electrons only, as in the example shown, there are cases in which it accelerates the positive ions more than it does the electrons. The momentum acquired from the force is so rapidly transferred to the ions that the force can be regarded as acting on the system as a whole, rather than on the electrons alone. This intimate interaction between the electrons and the lattice has often been overlooked in elementary treatments of the electron theory of metals, but it is required by the observed facts, it finds its explanation in the quantum mechanical treatment of the interaction between electrons and ions, and it is an important part of the general picture.

I have undertaken to outline the treatment of two points in which the application of quantum mechanics and the Fermi statistics has eliminated serious difficulties in the classical theory of metals. The outstanding problem still remaining is the explanation of superconductivity. No generally acceptable suggestion has yet been offered to account for the sudden complete disappearance of electrical resistance in some metals as they are cooled to very low temperatures. It is expected that such a suggestion will arise out of a more detailed and quantitative understanding of the interactions between the electrons and the ions, but until such an understanding is acquired, each feature of the theory needs critical reexamination.

ON THE NECESSITY FOR SPECTRAL SUB-CLASSIFICATION, WITH A DISCUSSION OF PRELIMINARY RESULTS FOR 1300 SOUTHERN STARS

DORRIT HOFFLEIT

Harvard College Observatory

(Communicated by Harlow Shapley)

ABSTRACT

The importance of the determination of spectral sub-classes as a preliminary step in the determination of spectroscopic absolute magnitudes is pointed out. Sub-classes have already been determined at Harvard for over 1700 southern stars of types A5 through M, and mostly brighter than the seventh visual magnitude (Table II). The criteria used are listed in Table I and are discussed briefly. Accidental errors in the sub-classes will be reflected in the determinations of absolute magnitude, increasing the observed dispersion in the latter for a given observed spectral sub-class. A comparison between the total observational error-dispersion and the "true" dispersions in the absolute magnitudes appears to indicate that the Harvard sub-classes are sufficiently accurate to justify the continuance of the absolute magnitude determinations.

The Harvard sub-classes are compared with the Henry Draper classes and with the sub-classes determined at Mount Wilson and at Bonn (Figs. 2–5). Various possible sources for small systematic differences are considered and it is found that the use of somewhat poorly exposed plates may become an important source of error.

1. *The Necessity for Sub-Classes.* When the Henry Draper system of stellar classification was first adopted no very fine subdivision of the classes was attempted. The small scale of most of the plates used for the Draper Catalogue and the Extension made a closer subdivision hardly possible, and for a great many purposes it is unnecessary. From the calibration curves for the intensities of the lines used in the determination of spectroscopic absolute magnitudes it has become evident, however, that a more detailed system is really needed. For example, a diagram published by Young and Harper[1] indicates that the absolute magnitude corresponding to one specific value of the ratio of the intensities in the pair λ 4071 and λ 4077 varies *four* magnitudes from spectral class F8 to G8.

Fig. 1 is plotted from tabular data on calibration curves given by the Mount Wilson observers[2] for the intensity ratio of the lines 4216: 4250 (SrII : FeI). From the steepness of the curves it is at once appar-

[1] *Pub. D. A. O.*, **3**, 18, Fig. 7, 1924.

[2] *Mt. W. Contr.* 199, 1921 (*Ap. J.*, **53**, 13, 1921); the more recent Mt. W. Catalogue, *Mt. W. Contr.* 511, 1935 (*Ap. J.*, **81**, 187, 1935), does not give calibration curves or data on line intensities.

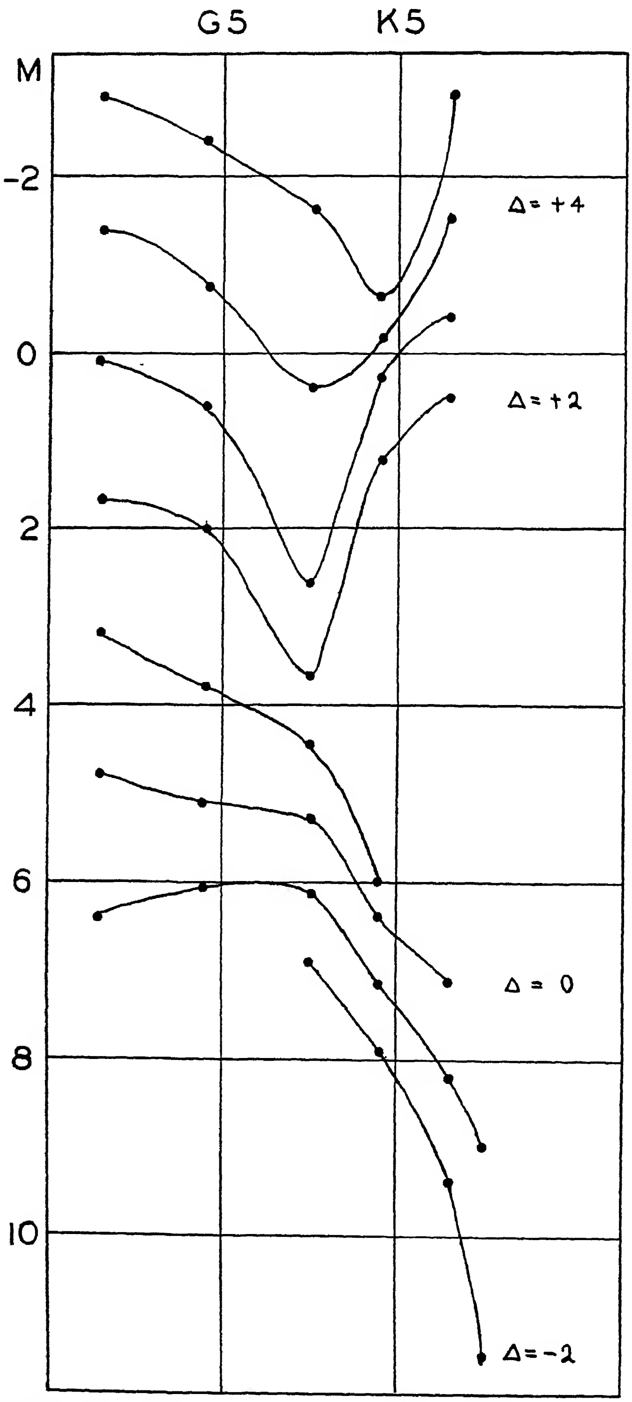


FIG. 1. Calibration curves for λλ 4215 : 4250 (after Mount Wilson Contr., No. 199, 1921). Curves represent constant values of the intensity ratios.

ent that a knowledge of the sub-classes is necessary in order to determine the absolute magnitudes of stars from G0 to M. The data given in Mount Wilson Contribution 199 are insufficient to show the trend of the curves for earlier spectral classes. A graph illustrating the dependence on spectral class of the ratios for three absolute-magnitude pairs was, however, given in an earlier paper by Adams [3]; it likewise indicates a rapid variation for the late spectral types, but comparatively little variation with spectral class for types earlier than G0. Nevertheless preliminary Harvard results indicate that even among the earlier classes the variation of the strength of absolute-magnitude lines with spectral class cannot be neglected.

2. The Criteria. The criteria I have used in the sub-classification of spectra are given in Table I. They are principally the final Draper

TABLE I

HARVARD CRITERIA FOR SPECTRAL CLASS

A5	$H + H\epsilon = 0.9K$ $K > H\epsilon$ metallic lines increasing	G5	$H\gamma < 4325$
F0	$H + H\epsilon = K$ $K = 3.0H\delta$	K0	$H\gamma = 0.5$Fe 4325 4227 three times as strong as in G0 H and K at max. 4227 = 2FeII 4172 4227 = 3FeI 4383 G band stronger than 4227
F5	Fe 4325 $= 0.1H\gamma$ G band 4227 $= 0.5H\gamma$	K5	H, K, 4227 most conspicuous. G band not continuous. 4383, 4405 prominent.
G0	$H\gamma = 1.5$Fe 4325 $H\delta = 4227$ H and K very strong G band well defined	M	TiO bands increasing. Spectrum faint in violet.

criteria given by Miss Cannon in the *Transactions of the International Astronomical Union* for 1935.[4] As my method consisted in a line-for-line comparison of spectra placed adjacent to one another and the subsequent interpolation of a type between two standards, I paid less attention to the actual values given by Miss Cannon for the intensities or intensity-ratios of certain critical lines than to the progression of such values toward the later spectral types.[5]

The "multiple nature" of the Harvard criteria has frequently been commented upon.[6] For my part I have experienced uncertainty be-

[3] *Mt. W. Comm.* 26, 22, 1916.

[4] Vol. 5, pp. 181–183.

[5] Graphs illustrating the variations of some of the lines used for classification are given by Menzel in *H.C.* 258, 1924; *cf.* also, C. H. Payne, *Stellar Atmospheres*, 121 ff., 1925; Curtiss, *Handb. d. Ap.*, V_1, 54, 1932.

[6] *cf.* Young and Harper, *op. cit.*, p. 14.

cause of conflicting results from the various criteria mainly in the A and the K types. Among the A stars the metallic lines and the ratio H to K frequently indicate different sub-classes, amounting in extreme cases to the difference from A4 to F4. As the metallic lines are too faint to have been seen in the majority of the early-type stars classified in the Draper Catalogue, I have felt it to be most consistent with the system of that catalogue to neglect the intensities of the metallic lines whenever the two criteria appear contradictory, and to adopt the sub-class indicated by the stronger H and K lines.

Among the K-type stars the ratio of Hγ to 4325 Fe and the intensity of 4227 are the principal criteria. The hydrogen intensity seems to place numerous stars in sub-class G5 whereas 4227 indicates that the sub-class should be K2. The ratio of Hγ to 4325 is less sensitive than the intensity of 4227 to change in class at K0 to K2, but unfortunately both criteria are subject to absolute-magnitude effects. When the two criteria appeared contradictory, I used the absolute strengths of the lines 4383 and 4405 as the deciding factors.

For the M-type stars Miss Cannon used the width of 4227 as a criterion auxiliary to the TiO bands. On most of the large-dispersion Harvard plates the continuous spectrum at λ 4227 is already too faint to make the width of this line a practical criterion. I have therefore classified only on the basis of the titanium-oxide bands.

Miss Cannon mentions the line 4077 SrII among the criteria for the G-type stars. As the intensity of the line is one of the most important absolute-magnitude criteria, I have not used it at all for classification purposes.

TABLE II

THE MATERIAL CLASSIFIED

Spectral Class (H.D.):	A	F	G	K	M	Total	H.R. not Classified
]5^m	19	73	108	144	31	375	2
5^m–6^m	38	89	70	314	42	553	136
6^m–7^m	40	91	106	285	27	549	535
[7^m	23	53	57	110	19	262	
Totals	120	306	341	853	119	1739	673

3. The Material. A summary of the numbers of stars that have been sub-classified is given in Table II. Preparatory to the undertaking, a card catalogue, complete to $6^m.5$ (visual), had been compiled from the

Revised Harvard Photometry [7] (hereafter called the H.R. Catalogue) for all the stars south of $-20°$ the spectral classes of which according to the Draper Catalogue are A5 or later. Variable stars and stars with composite spectra were not included. Plates made with the 13-inch Boyden telescope are not yet available for all of the H.R. stars (see table); but many of the plates exposed on brighter stars show also spectra of good quality for stars fainter than those listed in the H.R. Catalogue. Accordingly all the usable spectra of faint stars that happened to occur on these plates were sub-classified—a total of over 1700 stars. At least one adequate plate was available for 65 per cent (1296 stars) of the 1969 H.R. stars. Good plates were available for all but two stars brighter than 5^m, for 80 per cent of the stars between 5^m and 6^m, and for 41 per cent of the stars fainter than 6^m that are included in the H.R. Catalogue. Of the stars fainter than 5^m, however, more than one plate was available for only 22 per cent (540 stars).

The actual tabulation of the new spectral classes is not given here, as it will form a part of the larger table of the spectroscopic absolute magnitudes that will be ready for publication within a few years.

4. The Internal Accuracy of the Estimates of Sub-Class. It is of importance to determine the accuracy of the new sub-classes; for if the dispersion in classification is sufficiently great it alone will produce a spurious dispersion in the absolute magnitudes.

Let σ_T represent the total observed dispersion in absolute magnitude for a given observed spectral sub-class (giants and dwarfs being treated separately); σ_M, the dispersion in absolute magnitude arising from errors in the estimates of line-intensities and from discrepancies between the absolute magnitudes derived from the various pairs of lines used; and σ_S, the dispersion in the estimates of sub-class.

If λ is the slope of one branch of a Russell diagram (considered linear over a short range in spectrum), then the dispersion in the absolute-magnitude coordinate arising from the dispersion in the spectral-class coordinate is $\lambda\sigma_S$. Unless

$$\sigma_S < \frac{1}{\lambda}\sqrt{\sigma_T^2 - \sigma_M^2},$$

the observed range in the absolute magnitudes is entirely the consequence of observational errors and therefore meaningless.

In Table III I have listed, for various spectral classes and apparent magnitude intervals, the average differences in sub-class estimated from two plates. There appears to be no specific dependence of the errors

[7] *H.A.*, 50, 1908.

either on class or on apparent magnitude. Although the tabulated values were not derived from all of the material classified, the selection of data was entirely random. The estimates for one hundred and fifty representative bright stars indicate that the dispersion, σ_S, is 0.8 (corresponding to a probable error for one estimate of $\pm$ 0.5 sub-class).

TABLE III

AVERAGE DIFFERENCE IN ESTIMATES FROM TWO PLATES (IN TENTHS OF A SPECTRAL CLASS)

Spectral Class:	A	F	G	K	M	Weighted Mean
]5^m	0.5	1.1	1.3	1.3	1.1	1.2
6^m–7^m	1.3	1.7	1.3	1.2	1.1	1.35
[7^m	—	1.1	1.2	1.5	0.5	1.2

The error-dispersion in absolute magnitude, σ_M, for the bright stars of Harvard Annals, **105**, Number 3, was found to be $0^M.22$, and the value for λ approximately 2/17 for G and K giants.

For σ_T, I have substituted a value of the "true" absolute-magnitude dispersion. Although the "true" is less than the total observed dispersion, determinations of absolute magnitude will be of little significance unless the above inequality holds as well for the "true" dispersion. Schalén [8] found for giant G5 to K2 stars a true dispersion of $0^M.9$; and Gyllenberg,[9] through an analysis of the apparent-magnitude and proper-motion distributions of A and F type stars, found a "true" dispersion of the order of $0^M.8$. Gyllenberg remarked, however, that the dispersion in absolute magnitude for stars brighter than a limiting *apparent* magnitude (the stars for which spectroscopic absolute magnitudes were determined were brighter than $5^m.1$) is always less than the true dispersion for *all* the stars. For absolute magnitudes computed from trigonometric parallaxes, Nassau and McCuskey [10] have found a "true" absolute-magnitude dispersion of the order of $1^M.5$ for each spectral class. I have adopted $0^M.8$ as a conservative estimate. The inequality

$$\sigma_S < \frac{1}{\lambda}\sqrt{\sigma_T^2 - \sigma_M^2}$$

or

$$0.8 < \frac{17}{2}\sqrt{0.64 - 0.05}$$
$$< 6.5$$

[8] *Upsala Medd.* 55, 1931.
[9] *Lund Circular* 12, 1937.
[10] *M.N.*, 94, 564, 1934.

is thus satisfied, and it appears that the errors in the sub-classes are not sufficient to invalidate subsequent absolute-magnitude determinations for the giant stars.

Hopmann [11] has recently found that the true dispersion in absolute magnitude for the main-sequence stars is only $0^{M}.4$. For this value the above inequality would be much less pronounced ($0.8 < 2.4$), indicating the necessity for a higher degree of precision in the sub-classification of dwarf than of giant stars.

5. Comparison of Revised Harvard Classes with Other Determinations. There are three classifications with which the sub-classes determined for the southern stars may be compared: (1) the original Draper classes; (2) the classes determined at the Mount Wilson Observatory for stars in an overlapping region accessible to both northern and southern instruments; and (3) classifications from small-dispersion plates by Becker and Kohlschütter [12] for about one hundred stars in Selected Areas.

A comparison between the Harvard sub-classes and other systems for stars brighter than 5^{m} has already been published in the Harvard Annals.[13] It indicated a slight difference in scale between the Harvard and Mount Wilson systems such that Harvard classes F4 to M0 correspond to Mount Wilson F3 to M1. Fig. 2 of the present paper shows the correlations between the revised (abscissæ) and the Draper classes (ordinates) for the three magnitude groups: (*a*) 5^{m} to 6^{m}, (*b*) 6^{m} to 7^{m}, and (*c*) fainter than 7^{m}. For the sub-classification of all of these stars the comparison plates used were those that had been employed for the brighter stars. Within the accidental errors, the correlations are the same as the correlations found for the bright stars except at G5. The faint Draper G5 stars have on the average been sub-classified as of later types. One hundred and seven of the stars fainter than 5^{m} were also classified at Mount Wilson. The comparison (Fig. 3) is not systematically different from the corresponding comparison for the brighter stars.[13] For this reason an investigation of the possibility of any systematic variation in the Draper system of classification seemed advisable. Systematic differences might conceivably exist between the classes that had been estimated from small and from large-scale plates. In fact, Humason's classifications in Selected Areas indicate such an effect in the Mount Wilson material:[14] his plot of the spectral classes estimated for 90 bright stars on small dispersion (114 A/mm) slitless and on large dispersion (30 A/mm) slit spectra shows that there was a

[11] *Berichten d. Math.-Phys. Kl. d. Sächsischen Ak. d. Wissen. z. Leipzig*, **90**, 204, 1938.
[12] *Bonn Veröff.* 27, 1933; 29, 1936.
[13] *H A.*, **105**, 45, 1937.
[14] *Mt. W. Contr.* 458; *Ap. J.*, **76**, 224, 1932.

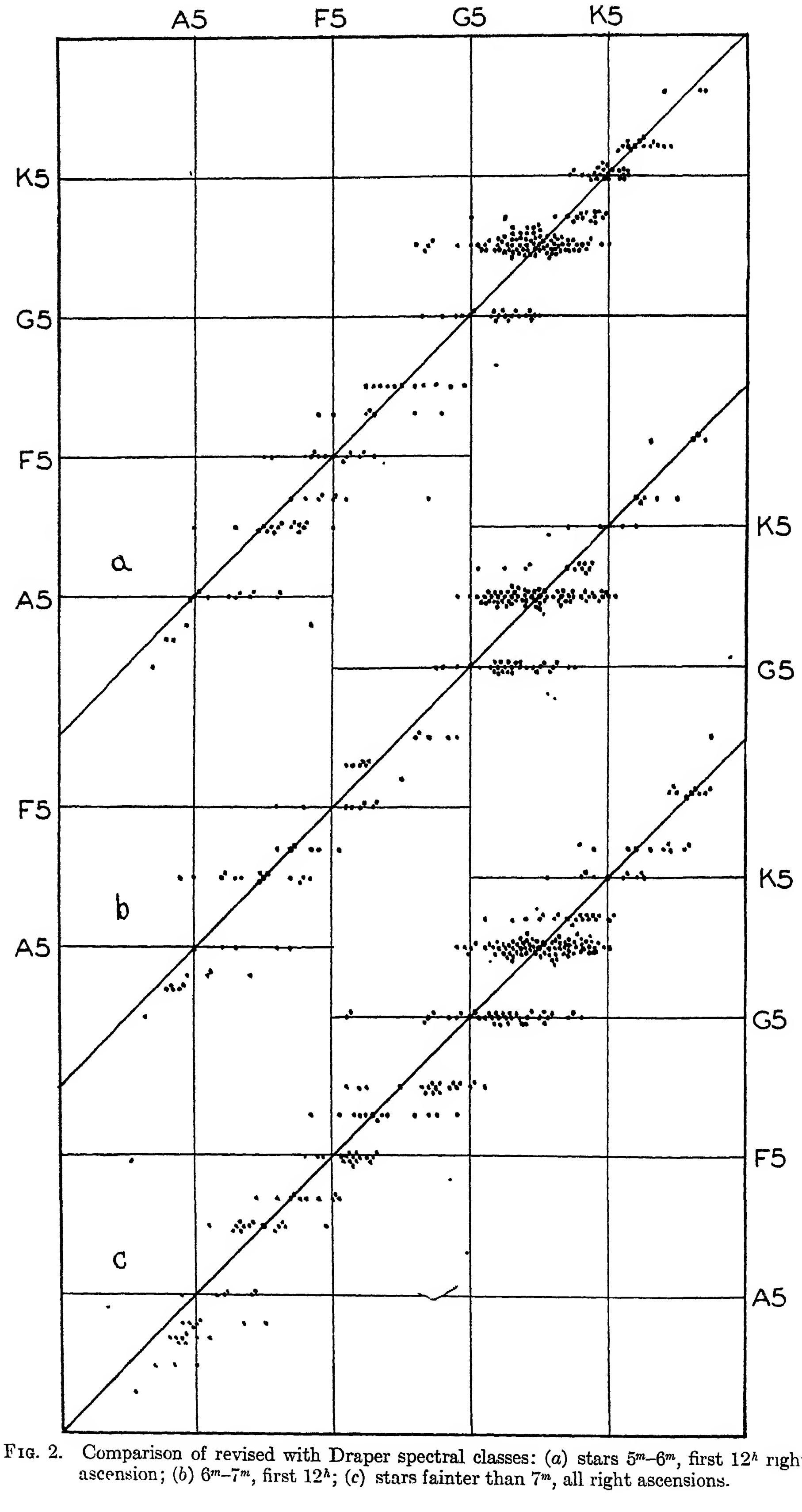


Fig. 2. Comparison of revised with Draper spectral classes: (*a*) stars 5^m–6^m, first 12^h right ascension; (*b*) 6^m–7^m, first 12^h; (*c*) stars fainter than 7^m, all right ascensions.

tendency to estimate the A stars systematically later and the G and K stars systematically earlier on small than on large dispersion plates.

For the Harvard material I ascertained from the last column of the Henry Draper Catalogue [15] which of the stars in the first twelve hours of

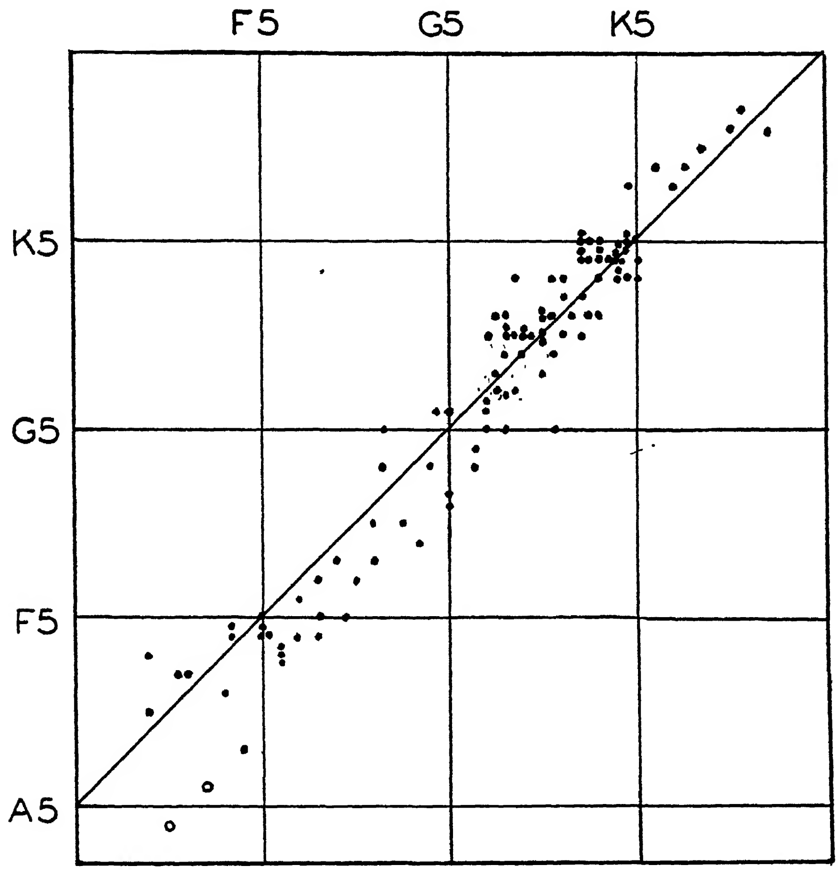


Fig. 3. Comparison between revised Harvard classes (abscissæ) and Mount Wilson classes. Circles denote spectra designated "*n*" at Mount Wilson.

right ascension Miss Cannon had classified from B-plates (8-inch Bache camera, dispersion 400 A/mm), and which from X-plates (the series used in this investigation, dispersion 45 A/mm). Fig. 4, a plot of the present sub-classes against Miss Cannon's classes (*a*) from small and (*b*) from large dispersion spectra, reveals no obvious systematic difference. In Fig. 4c I have plotted only those of my sub-classes that were determined from two plates of good quality yielding results accordant within 0.2

[15] *H.A.*, 90–100.

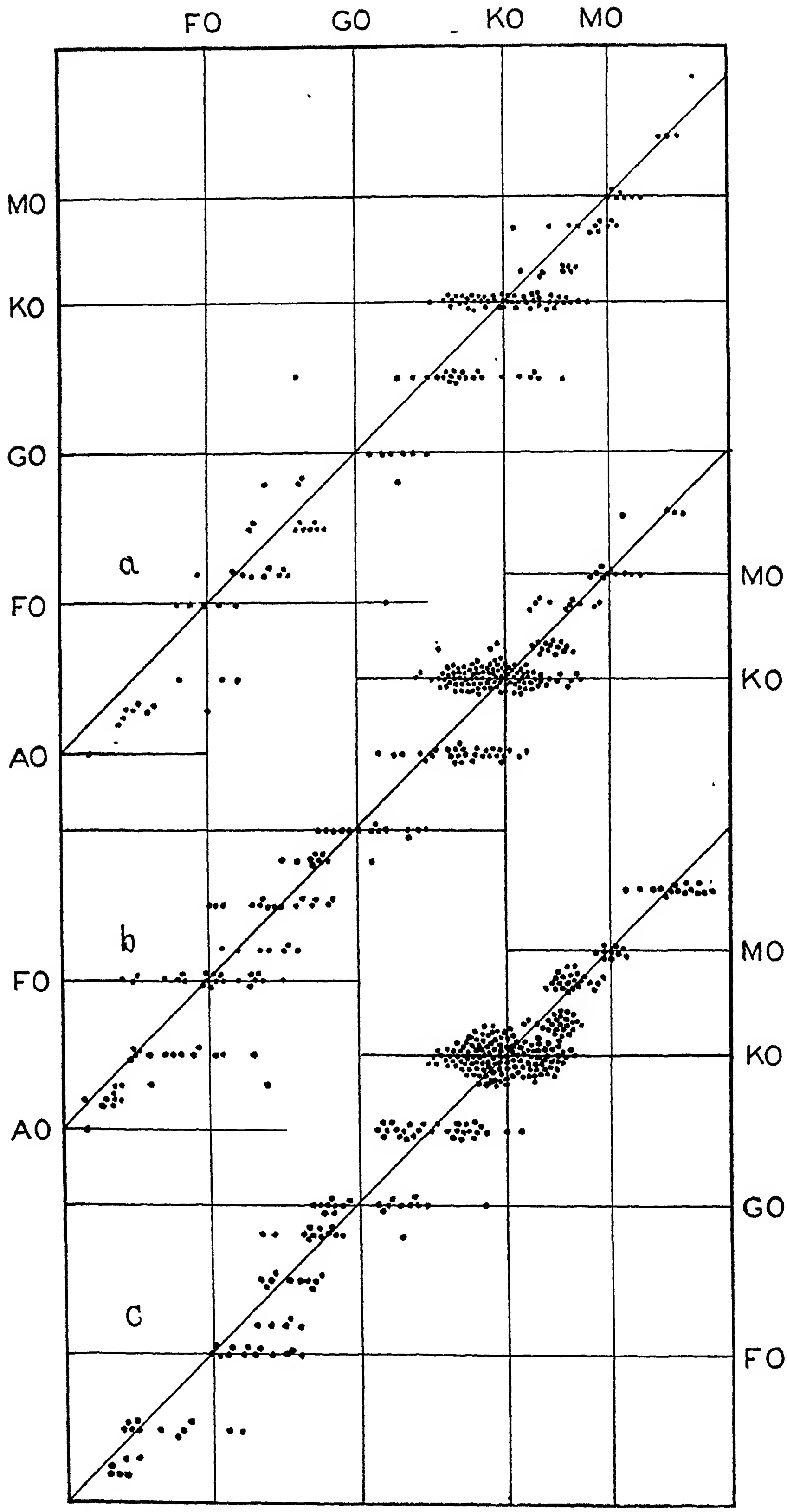


FIG. 4. Comparison of revised Harvard classes with (*a*) classes determined by Miss Cannon from small dispersion, (*b*) from large dispersion, spectra; (*c*) results from only first class plate material.

classes. The discrepancy between the bright and the faint stars at G5 is no longer conspicuous: it is therefore chiefly dependent on the quality of the plate material, and strongly suggests caution in the use of poorly exposed plates.

A comparison of the Harvard sub-classes with those determined by Becker and Kohlschütter[12] (Fig. 5) indicates that the scales of the two

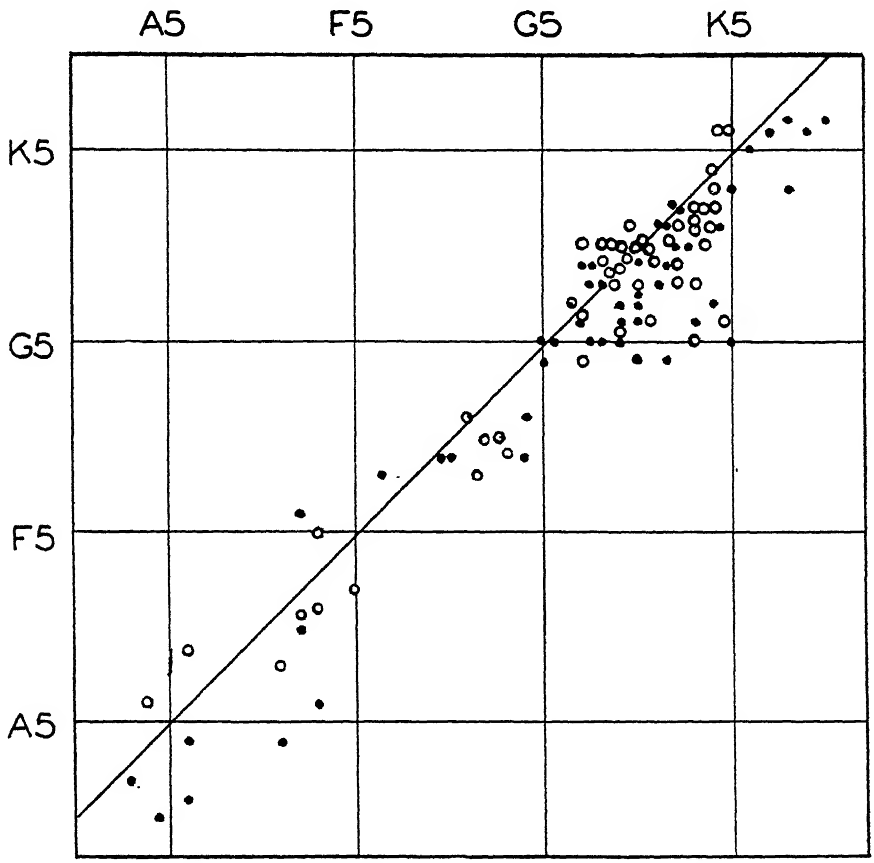


Fig. 5. Comparison of revised Harvard classes (abscissæ) with those determined by Becker and Kohlschütter (ordinates). Dots, Bonn Veröff., No. 27, 1933; circles, *ibid.*, No. 29, 1936.

systems are the same; but that Becker and Kohlschütter classify systematically 0.2 classes earlier. The smallness of the scattering suggests that lower dispersion than I have employed is useful for sub-classification.

6. Frequencies of Spectral Classes. The frequencies of the spectral classes of the brightest and of the faintest stars studied are shown in Fig.

6. The full-drawn curve represents the frequencies for stars brighter than 5^m; the broken curve, the frequency for stars fainter than 7^m. The minimum at G0 for the former may be attributed to the selection of apparently bright stars only. By limiting the investigation to the

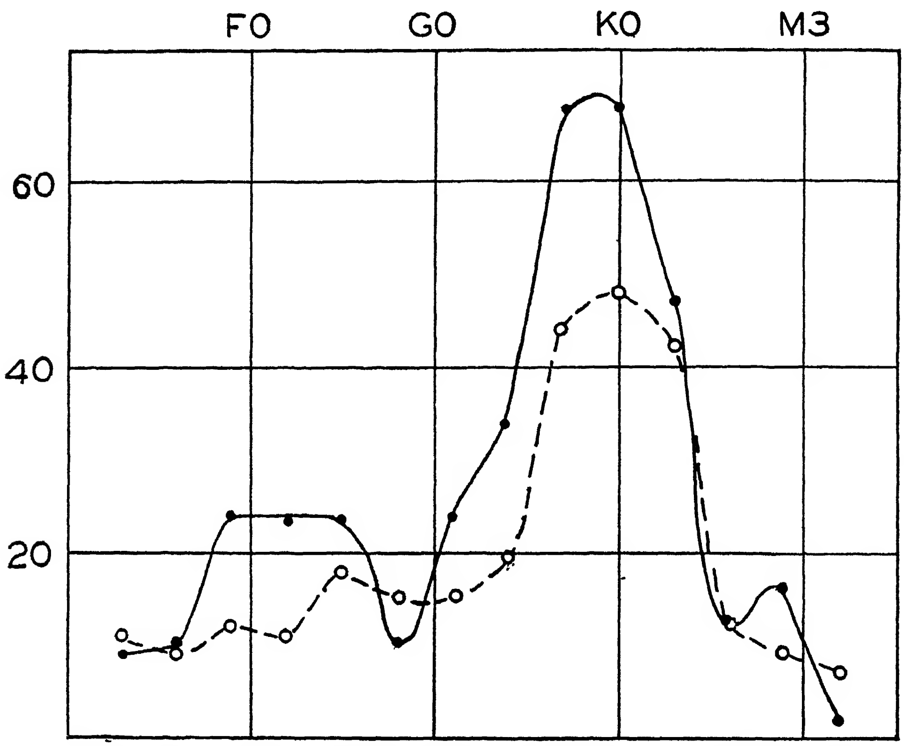


Fig. 6. Frequencies of spectral classes: dots, stars brighter than 5^m; circles, fainter than 7^m.

naked-eye stars we necessarily include very few stars fainter than absolute magnitude + 3; the main sequence G stars have somewhat fainter absolute magnitudes and the giant stars of classes F5–G5 are cosmically infrequent—hence the minimum at G0. The main sequence stars of the earlier classes are sufficiently luminous to have been included in the list of apparently bright stars; this accounts for the secondary maximum at F3. Among the apparently faint stars the proportion of dwarfs increases. Thus for stars with apparent magnitudes between 7^m and 8^m the minimum at G0 has disappeared, as the main sequence G stars outnumber the more luminous and more distant A stars.

SUMMARY

1. Because of the gaps that occur in the Henry Draper system the spectral classes given in the Draper Catalogue have generally been found insufficiently subdivided for determinations of spectroscopic absolute magnitudes. The necessity for more detailed sub-classification is discussed.

2. The criteria used for sub-classification in the present study are listed and the multiple nature of some of them are briefly considered.

3. The effect of errors in the spectral sub-classes upon the subsequent determination of absolute magnitudes is described. For stars of 5^m to 7^m the errors in the sub-classes newly determined from Harvard large-dispersion spectra are sufficiently small to justify use of the sub-classes in the determination of absolute magnitudes.

4. The new sub-classes for over 1300 southern stars fainter than the fifth magnitude are compared with the Draper classes and with sub-classes determined at Mount Wilson and at Bonn. Possible effects dependent on dispersion and plate quality are considered. Poorly exposed plates are apparently an important source of systematic error in sub-classification.

GEOLOGY AND PALEONTOLOGY OF LAKE TACARIGUA, VENEZUELA

EDWARD W. BERRY

Professor of Paleontology, Johns Hopkins University

(Read by title, April 20, 1939)

ABSTRACT

An account of the results of a reconnaisance of the geology of the Tacarigua basin in northern Venezuela with particular attention to the system of elevated lake terraces. This region has been inhabited for untold centuries and the terraces abound in archeological and anthropological objects. Evidence will be presented to show that the lake terraces are post-Columbian in age.

INTRODUCTION

IN 1933 I had the pleasure of spending 3 months in Venezuela as the guest of Dr. Rafael Requena, at that time Secretary to General Gomez, and one of the most cultured gentlemen that it has been my good fortune to have as a friend. Dr. Requena found time in spite of onerous official duties to cultivate an absorbing interest in the archeological riches of his country.

The purpose of my visit was a study of the geology and paleontology of the deposits which were yielding such a quantity of archeological and anthropological objects. Unfortunately these studies were terminated by circumstances beyond my control after systematic stratigraphic work had reached a depth of but 3 meters and the deeper and more interesting levels had not been reached. In spite of this it seems worth while to publish some account of what was accomplished.

I have to acknowledge my indebtedness to Dr. Requena and the Venezuelan government, to my good friend Dr. James A. Tong who was with me throughout the campaign and to whom any success which it had was due, and to Alfred Kidder II, who worked with us during part of the time.

The following gentlemen have examined material for me, and I wish to record my grateful appreciation for their assistance: Tow-net samples from the Lake were examined by Professor Chancey Juday, limnologist of the Geological and Natural History Survey of Wisconsin; Terrestrial and fresh-water Mollusca by Dr. Henry A. Pilsbry of the Academy of Natural Sciences of Philadelphia and Professor Frank C. Baker of the University of Illinois; Fish bones by Dr. George S. Myers

of the U. S. National Museum; Bird bones by Dr. Alexander Wetmore, Assistant Secretary of the Smithsonian Institution; Diatoms by K. E. Lohman of the U. S. Geological Survey; Mammal bones by Gerrit S. Miller, Jr., and Remington Kellogg of the U. S. National Museum; Ostracods by Dr. C. H. Blake of the Massachusetts Institute of Technology; and calcareous algæ by Dr. Marshall A. Howe, late Director of the N. Y. Botanical Garden.

Little has been written about the immediate region. Aside from Codazzi's classic publication on the natural history of the general region of New Granada the following publications have come to my attention:

Codazzi, A., *Jeografía fisica y política de los Estados Unidos de Colombia,* herausg. von Felipe Pérez. 2 vols. Bogota 1862–63.

Marcano, G., *Ethnographíe précolombienne du Vénézuela, Vallées d'Aragua et de Caracas,* Paris 1889 (not seen).

Spinden, Herbert J., "New Data on the Archeology of Venezuela," *Proc. Nat. Acad. Sci.,* 2, 325–328, 1916.

Eigenmann, C. H., "The Fishes of Lake Valencia, Caracas and the Rio Tuy at El Concejo," *Indiana University Studies,* No. 7, 1920.

Requena, Rafael, *Vestigios de la Atlántida,* Caracas, 1932.

Jahn, A., "Los craneos deformados de los Aborigenes de los valles de Aragua," *Soc. Venez. de Ciencías Naturales de Caracas,* Bol. 8, pp. 1–14, figs. 1–4, 1932.

Pearse, A. S., "Crustacea from Lake Valencia, Venezuela," *U. S. Natl. Mus. Proc.,* 59, 459–462, 1921.

—— "The Fishes of Lake Valencia, Venezuela," *Univ. of Wis. Studies in Science,* No. 1, 1920.

Berry, Charles T., "Pleistocene Remains Found near Lake Tacarigua, Venezuela," *Journ. Wash. Acad. Sci.,* 24, 387–395, 500, 1934.

Kidder, A. V., 2nd, "The Harvard Archeological Expedition to Venezuela," *Sci.,* 81, 222, 1935.

Wetmore, Alex., "Pre-Columbian Bird Remains from Venezuela," *The Auk,* 52, July, 1935.

Geography

Lake Tacarigua, or Valencia, as it is more commonly called, lies in what appears to be a structural basin in the Coast Range of Venezuela, about 25 km. in an air line from the Caribbean. It has a surpassingly beautiful setting and during the over a quarter of a century of his dictatorship, the region was the preferred residence of General Gomez, and in his later years he never left it except for annual visits to the coast or to the baths at San Juan de los Morros.

The Divide between the Caribbean and the Lake in this longitude rises to culminating peaks of from 6000 to 7600 ft. (Candelaria). The southern rampart of the Aragua basin, as it is usually called, is a less rugged terrain. Los Morros which separates the Coast Range district from the Orinoco plain has an altitude at San Juan of 4706 ft.

The Mountains increase in altitude to the eastward and have their highest peaks just east of Caracas, and decline toward the west.

The Lake is roughly elliptical in outline, with the maximum diameter of about 85 km. parallel with the trend of the mountains, here east and west, and a width of about 25 km. Its area (planimeter measurements from the Jahn map, the most accurate of the region) is 436 sq. kms. Filling as it does the major part of the width of the Aragua basin, the mountains come down to within 4 or 5 kms. of the north and south shores, whereas to the eastward low fertile plains extend as far as San Mateo, the historic hacienda of Simon Bolivar, between 15 and 20 kms. Westward similar plains with a gradually diminishing rainfall extend indefinitely and merge with the plain of Carabobo about 14 km. west of Valencia.

A mountain spur from the north swings down to the Lake at La Cabrera, just 135 km. from Caracas. The railroad and cart road (Carretera), the latter a cement auto road, cross this spur at the pass of same name, famous as a strategic point in the War of Independence and in the later civil strife between the caudillos. The ridge continues its swing to the west and forms the axis of the peninsula of La Cabrera. Across the Lake south of La Cabrera is another upland peninsula Y-shaped, the two arms of the Y being Macapo ridge extending northeasterly and Yuma ridge extending northwesterly. Several islands in the Lake with metamorphic rock cores apparently represent the culmination of submerged ridges connecting the aforementioned peninsulas. Except for these two peninsulas the lake-shores are gently curved and bordered with lake terraces, the sites of banana plantations.

The peninsula of La Cabrera in general outline is a narrow isosceles triangle, with its narrow base to the east. The south shore is deeply indented due to the configuration of the basement rocks, giving the whole a suggestive resemblance to a Chinese dragon. The north shore, more protected from wave action, is straighter. The crystalline core rises steeply to heights of three or four hundred feet and is densely covered with vegetation. It is surrounded on all sides by a series of old lake terraces which dip slightly away from the core and toward the present lake. These constitute the arable land which is for the most part under cultivation and seems exceedingly fertile.

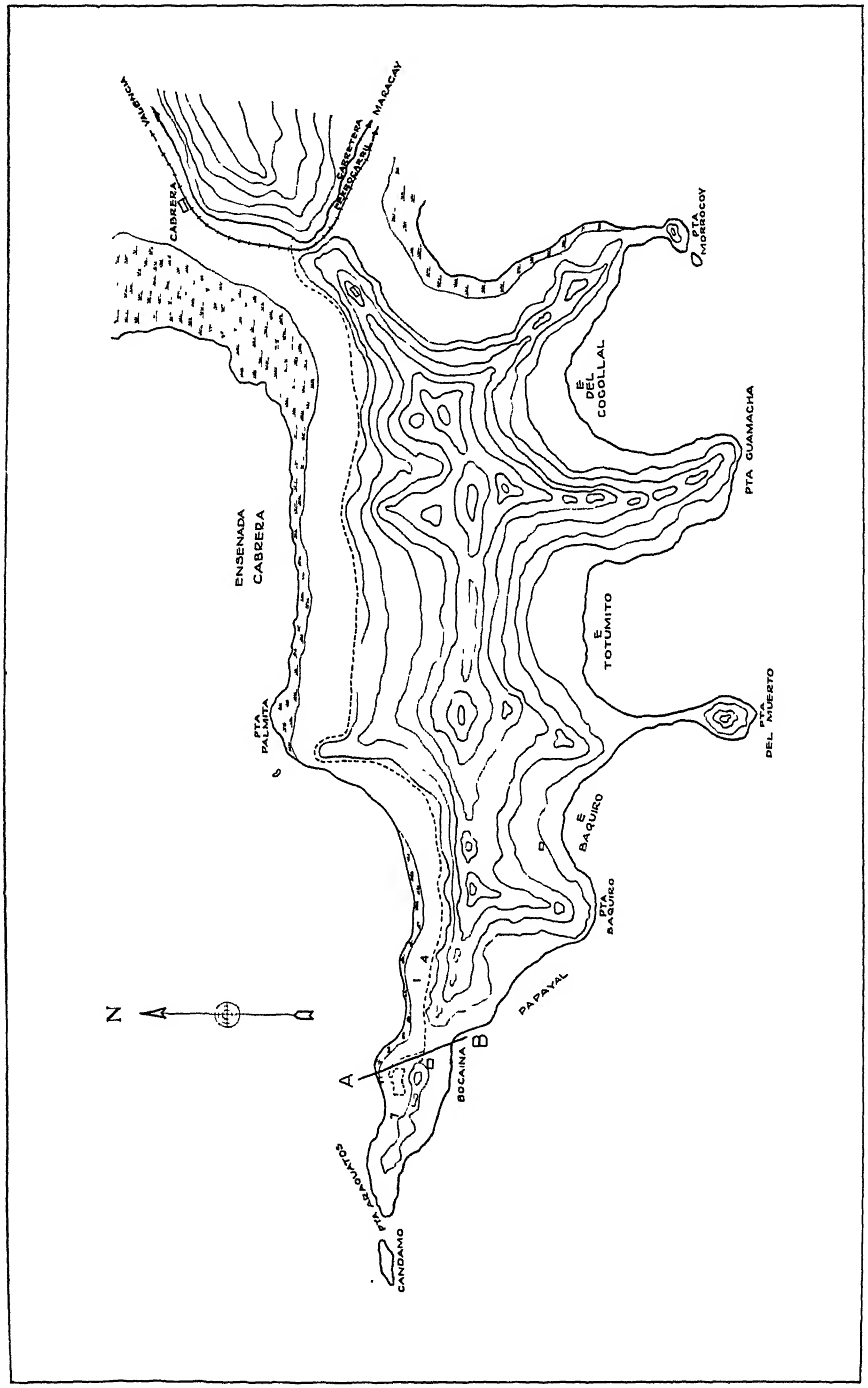


FIG. 1. Sketch map of La Cabrera Peninsula.

About a quarter of a mile from the western tip, there is a saddle in the basement rocks which rise again to the west of the saddle to form the core of what must have been an island during the maximum submergence of the region. Whether because this was an island, or from its strategic position, easily defended from attacks by land or wild animals, it was apparently densely populated by the aborigines. Either this is true or they assembled here to bury their dead and to hold the ceremonial carousals accompanying the custom of secondary burials. (See map, Fig. 1.)

Although the alluvial terraces which surround the Lake and its islands are developed everywhere except on the most precipitous slopes, and supported a considerable population, as did also the flats at the east and west ends of the Lake and the evidence of human occupation is found everywhere around its perimeters, it seems concentrated on La Cabrera. Perhaps this is illusory and is due to the fact that over a considerable period there has been sporadic digging of small openings and trenches. Hence La Cabrera has quite a local reputation.

The only attempt which may be said to have been conceived in a scientific spirit was that of Dr. Requena and this was prosecuted almost entirely by subordinates, whose jobs depended on their pleasing their employer, and who possessed slight competence. The Requena excavations were extensive and a venerable tamarind tree gave its name to the site, which is generally referred to as Los Tamarindos or Tamarindo. From this site and others Dr. Requena built up an extensive collection of archeological objects and human remains, which culminated in 1932 in the publication of a book, *Vestigios de la Atlantida*, the main thesis of which was that Venezuela was first populated from the mythical continent of Atlantis which cataclysmally sank beneath the sea in Pleistocene times, and that these humans were the progenitors of the peoples whose various cultures we know as Mayan, Peruvian, and Mexican.

This hypothesis, coming from a man of such political eminence and one having intrinsically so romantic an appeal, was widely accepted in Venezuela and indeed attracted much attention throughout South America.

There have been 3 systematic archeological and anthropological studies of the Lake region. The one prosecuted by Alfred Kidder II in 1933 while with the writer, and supplemented in 1934, at the Tamarindo site. One by Wendell Bennett of the American Museum of Natural History in 1932, and another by Cornelius B. Osgood of the Peabody Museum of Yale University in 1933. The last two sites were

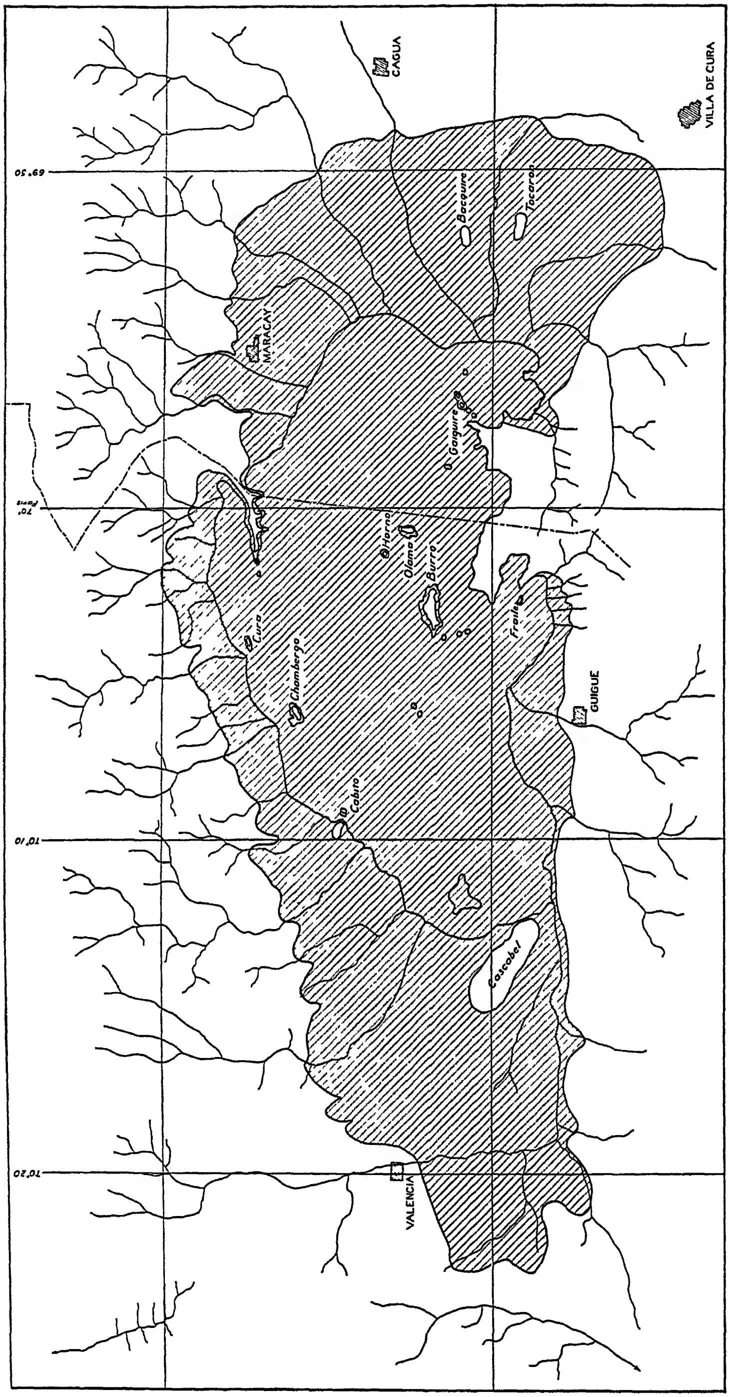


FIG. 2. Map of present and former extent of Lake Tacarigua.

the so-called mounds which are so abundant on the flats east of the Lake on the Hacienda Tocoron. No accounts of any of these have as yet been published.

Geology

The general geological setting is that of surrounding mountains made up of metamorphic and intrusive rocks. The basin is floored by fluviatile and lacustrine sediments the total thickness of which is unknown but is in excess of 450 ft., since of late years an artesian horizon has been exploited at about this depth. Its source is a sand bed, and no wells have struck bed rock. A Mr. Berk, who was in charge of the drilling at the time of my visit, reported a fossiliferous horizon from a depth of 180 ft. in a well at Tocoron, but this was based entirely on memory, and a visit to the site and an examination of the dumps was without results, so that although the statement is perfectly creditable we were unable to substantiate it by direct evidence. Unquestionably there should be numerous fossil horizons in the 450–500 feet of basin deposits.

Chronologically these basin deposits cover the time extending from some date in the Pliocene through the Pleistocene to the Recent. This rests on the opinion, derived from orogenic studies of the Andes, that the mountains which formed the Tacarigua or Aragua basin were elevated at some time during the Pliocene. I know of no way of getting at the true facts short of detailed studies of non-available well cuttings or long continued field studies, and both are precluded by the temperament of the people and the law of diminishing returns.

The latest geological episode in the history of the Lake is clear. At some time in the past it was much larger than it is at present. The accompanying map (Fig. 2) of this greater lake shows that its area of maximum extent was 1050 sq. kms. or nearly $2\frac{1}{2}$ times that of the present lake. Obviously nothing can be said about the fluctuations of level prior to the time of maximum submergence as this event destroyed any physiographic evidence that may have existed. In the same way, although it is reasonable to suppose that the shrinkage from this high level was not a continuous and regularly progressive process, but a fluctuating one, physiographic evidence of such fluctuations is destroyed and for the same reason.

There remains the record of five wave-washed terrace plains and their accompanying wave-cut scarps, the four oldest of which are emerged and face the lake at the edges of the successively lower levels, whereas the fifth is in process of formation by the waves of the present

lake along the shore where the lowest terrace plain dips beneath the waters of the present lake. This series of terraces line the shores in an imbricating series. The oldest and highest of these terraces at its inner (landward) margin lies against the much higher metamorphic rock cores of the surrounding hills and islands. At no place was this level observed in other situations, although there should be such exposures in the unexplored (by us) areas east and west of the Lake. A competent topographer would probably know all the answers after a single season's work. The surface of this highest observed terrace level shows but slight evidence of erosion and is 50 ft. above the present lake-level.[1] The surfaces slope gradually lakeward and on the island of El Horno where measured was 50 feet in width and its outer (lakeward) edge at the top of the scarp was 5 feet lower than its inner edge where it abutted the metamorphic core. The scarp is well developed, shows slight evidence of dissection and drops 13 feet to the inner edge of the next lower terrace level abutting against the older and higher terrace. This second plain is about 20 feet wide on El Horno and similarly its lakeward border is marked by a scarp with a drop of 11 feet. The third level is 15–20 feet in width on El Horno and its lakeward border is marked by a scarp which drops 10 feet to the 4th or next lower level, which is rather narrow on El Horno, not over 8 or 10 feet, with a scarp dropping 10 feet 8 inches to the fifth or lowest level which is here about 20 feet wide and extends to the water's edge.

The 3d level, the scarp of which has been cut into the basement rock on El Horno, as is poorly shown in the accompanying view (Pl. 2, Fig. 2), as also the algæ incrusted boulders (Pl. 3, Fig. 2) lying on its surface, although only 15 to 20 feet wide on that island, represents considerable lapsed time, since it is this level that forms the present surface of the country over such a large area both east and west of the Lake. It is this level on which the lake-dwellings were built the debris of which constitutes the so-called mounds which are so numerous on the flats of Tocaron, two of which were excavated by Bennett and Osgood respectively.

Proving the fluctuations mentioned in a preceding paragraph, the lowest scarp on El Horno shows several levels where the water stood long enough to record its level in the face of the scarp. This is also shown in numerous sections where water-laid beds cover intercalated humus deposits of terrestrial origin, which prove that minor advances

[1] Heights were determined by lock-levelling, but a modifying factor is that our work was carried out during the rainy season of 1933, and we were told by the captain of a banana boat plying on the lake that high water that season was higher than at any time since 1906.

of the shore line accompanied the temporary rising of the lake level at various times in the terrace record of subsidence from highest to lowest.

Probably the most important of these minor events, faintly recorded in the lowest scarp on El Horno, is the well defined lowest level along the south shore of La Cabrera peninsula where the full sweep of wave action from across the lake was brought into play. Here at Bocaina the lowest level extends 40 ft. back from the present shore, rising about 2 ft. in this distance. At its landward margin there has been developed a well defined scarp 3 to 4 ft. high the crest of which is 5 feet 4 inches above lake-level. A second wave cut bench 54 feet in width rises gradually to the next older scarp the crest of which is 10 feet 8 inches above lake-level and which corresponds to the lowest well developed scarp on El Horno. This twofold nature of the lowest main scarp around the lake is shown graphically in the accompanying profile across the saddle of La Cabrera peninsula (Fig. 3).

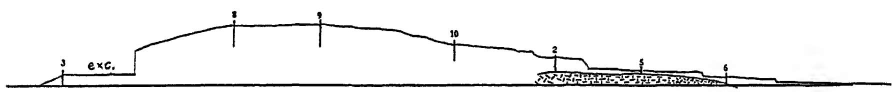


FIG. 3. Profile *A–B* La Cabrera Peninsula.

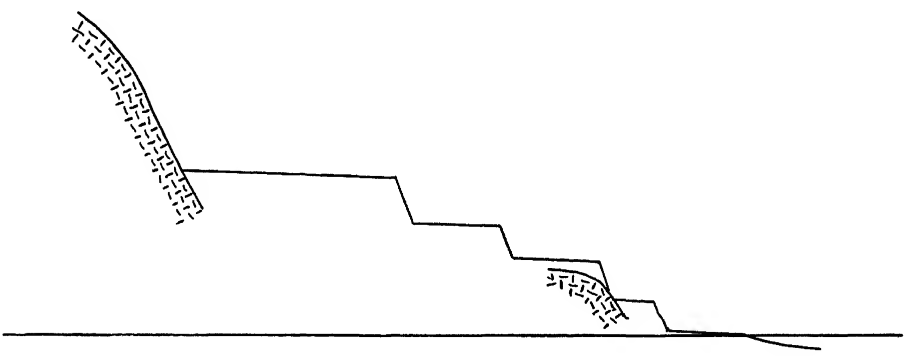


FIG. 4. Profile west half of Island El Horno.

The lake terraces described briefly above are particularly well shown around La Cabrera, especially in the present forest along Papayal and Ensenada Baquiro, around the islands, especially El Horno and Chambergo, on the south shore of the Lake at La Florida, and elsewhere (Pls. 1; 2; 3, Fig. 1).

The materials of which these terrace deposits are composed are fully described in the accompanying sections of dug pits, excavations, and natural exposures and scarcely warrant a separate treatment.

Where they were deposited at a distance from high land they comprise predominantly bluish clays and light diatomaceous Planorbis marls. Close to the uplands their materials include a varying amount of gravels, cobbles and boulders. These coarse beds, well displayed in the excavations at Tamarindo, were very puzzling when first seen. Their complete explanation is seen in the conditions at the present time in the forest wherever the alluvial deposits are in contact with the metamorphic cores of the hills. In such situations a talus of unassorted boulders covers the forest floor, sometimes as a definite talus at the base of the steeper declivities, more often as angular pieces of all sizes which have worked their way lakeward down the alluvial slopes. This would have been especially active in the past at those times that it was augmented by wave action.

If this situation can be visualized, and it is everywhere observable at the present time where the proper conditions exist, and must have been when the lake was at its maximum or whenever any of its stages washed the metamorphic rocks, it is easy to see how to-day one finds broken shells at the contacts between terrace materials and metamorphic basement rocks, or why in the sections, courses of heavy boulders may have pockets of Planorbis with unbroken shells, or there may be stringers of humus in lake sediments.

I may seem to labor this point, which seems so perfectly obvious. I do so because Señor Castillos, who was in charge during the Requena operations, was strongly of the opinion that these boulders, again exemplified in the entry at Tamarindo, were artificially placed and were the work of the aboriginal inhabitants of the site. There is not the slightest evidence that such was the case.

Age

The main problem which took me to Venezuela was the hope of determining the age of the human remains. It was hoped that the stratigraphic collection and identification of the associated fossils would solve this problem. Unfortunately this did not prove to be the case, and for the following reasons: (1) The work was never completed, and was systematically prosecuted to a depth of but 3 meters. (2) A lack of systematic knowledge of the Recent fauna which is so necessary to check the fossils with, especially to disclose subspecies or any changes of geographical distribution which may have occurred. (3) Absence of adequate osteological materials in American Museums. Partially corroborative is the fact that the more advanced fossils all belong to still existing species, whereas among the molluscs, ostracods and di-

atoms there are several forms new to science. Here again the fact as to whether these are extinct or not awaits their discovery or absence in the Recent.

Among the commonest fossils at all sites are the bones of fishes, which in places literally pack the ashy muds. Dr. Myers, who examined my material, was unable to furnish me with any identifications because of the lack of osteological material of the Recent fishes of the region. Contrasted with this is the long list of bird bones with which Dr. Wetmore furnished, due in part certainly to the extensive collection of this class of material in the National Museum.

As I shall show in the following paragraphs, I believe that I have convincing evidence of the age of the Lake terraces, but it should be emphasized that this throws no direct light on the beginnings of human occupation of the region, for this antedates the oldest terrace, and was probably much older. The last statement rests on the facts that both human bones and archeological objects are fairly abundant in Lake sediments at depths below the present surface of the ground where they could not possibly have gotten before the Lake expanded to the highest terrace level. In other words, the aborigines dwelt on the land which was subsequently flooded during the rising of the Lake-waters. What follows, then, relates to the age of the terraces (Pl. 4, Fig. 2; Pl. 5, Fig. 1).

First impressions of the terrace system were of venerable antiquity, especially as in some instances the scarps had trenched the metamorphic rocks, and in places even the most youthful level supported a forest of large trees. On the other hand, the Lake is sufficiently large to develop strong wave action and the degree of dissection is not consonant with great antiquity, especially when the heavy rainfall is taken into account. To be sure there can be no chronological standard, as to either the formation of terraces or the time necessary to dissect them, but certainly I have seen abandoned fields in our Southern States much more eroded under a much less rainfall than even the oldest Tacarigua terrace, and the youngest Pleistocene terrace plains in our Middle Atlantic States, also under a very much less rainfall than the Aragua basin, are obviously much older in their physiographic age.

The key to the age of the Lake terraces was observed in the southeast corner of the main excavation at Tamarindo. Here at 1 meter below the surface there was a veritable "Potters field" of secondary burials (Pl. 4, Fig. 1). Associated with these burials were quantities of animal bones which are enumerated in the accompanying lists as coming from Locality 3. Included in these were bones of the domestic

dog, *Canis familiaris*, and domestic cattle, *Bos taurus*, showing that the Spainards had been in the country long enough for the Caribs, or whoever they were, to have obtained these two stocks which had been introduced from Europe. That this may not have been long after the arrival of Europeans in Venezuela is indicated by the fact that nowhere did we discover glass beads or metal objects of any kind, which, one would think, would normally spread among the natives much faster than dogs or cattle.

Above these burials just alluded to there is a thin undisturbed layer of sharp Lake-sand with Planorbis shells and water-worn pottery fragments, which in turn is covered by a few inches of humus. This means that subsequent to the burials the Lake advanced over the site, so that the highest and oldest lake terrace is certainly post-Columbian in age, and the whole terrace system from highest to lowest represents an interval of not more than four centuries. As will be seen in the accompanying lists of vertebrate remains collected, the domestic dog is associated with a large number of human burials. All of these are of the so-called Carib type and represent the last aboriginal culture in the region (Pl. 4, Fig. 1). That this "Carib" culture was not indigenous around Lake Tacarigua is obvious. It always occurs in the Humus beds close to the present surface or in obvious intrusions in the beds immediately beneath this surficial layer. Individuals of the older culture, whose remains are found in and beneath lake-beds, sometimes many feet below the surface (Pl. 4, Fig. 2), and who undoubtedly lived along the shores of the Lake for an unknown period of time before it expanded and whom we suppose to represent some branch of the so-called Arawak stock, represent an, as yet, unsolved problem in chronology.

As there doubtless will be published eventually a scientific description by Mr. Kidder of the archeological objects secured during our explorations I refrain from any comments on this aspect of the work except to call attention to the excellence of the zoomorphic ornamentation on the pottery and the abundance of corals in the deposits. The latter I believe were used for shredding manioc (cassava) and are an indication that this was one of the food staples along with corn, especially as the deposits contain numerous shaped stone disks about 5 inches in diameter and about 6/16ths of an inch in thickness which I suppose were used for baking.

One is tempted to become lyrical over the life of the original lake-dwellers with the two great tropical food crops—corn and cassava—

with an abundance of game and fish, also tobacco, as is evinced by the abundance of pipes (cochimbos) in the older levels.

As will be seen from the accompanying lists of vertebrates collected, deer bones of several species are invariably present as well as bones of the small caiman of the lake, another food animal, many turtles, varied with peccary, tapir, capybara and birds of several species.

Marine shells and corals of Caribbean species are common in the surficial deposits. Among them the following have been recognized: *Codakia orbicularis*, *Fissurella rosea*, *Fissurella nodosa*, *Cittarium pica*, *Nerita versicolor*, *Tectarius muricatus*, *Strombus gigas*, *Oliva reticularis*, *Lucina* sp., *Charonia tritonis* var. *atlantica*.

Several of these undoubtedly were used as food, but it is doubtful whether they were ever transported from the coast and consumed at the Lake sites.

Although it is but 25 km. in an air line from the coast to the Lake, it must have been at least three times that distance by trail and it is probable that whatever eating of mollusks was done, was on or near the coast and that the shells were preserved as a rough medium of exchange since there was obviously a demand for them for ornamental purposes, as they are variously perforated and smoothed for stringing.

Fragments of *Strombus* are common in the deposits and from this thick shell large beads the size of marbles were manufactured. Other shell beads are cylindrical and 1 to 1.75 cm. long and 3 to 4 mm. in diameter. Still others were of tiny perforated disks 3 mm. in diameter and 0.5 to 1.5 mm. thick. Others are doubly perforated rectangles about 0.75 by 1.5 cm. and about 1.5 mm. thick. The aborigines also frequently carved rather æsthetic doves (palomas) of shell.

I will merely mention a great variety of mostly cylindrical beads and pendants of vein quartz, rose quartz, serpentine or nephrite. Olivas are perforated lengthwise. The Neritas and Tectarius were perforated through the body whorl and the Charonias were modelled into trumpets or horns for augmenting the voice, either for calling or ceremonial howling.

Musa (?) sp. (Pl. 7)

Three fragments which fit together and show that the leaf was more than 28 cm. in width. The slightly ferruginized impression of this large leaf forms one side of the flat objects which are from 1 to 3 cm. in thickness and consist of sandy clay which is partially indurated and iron stained especially toward the surface opposite the impression. The material has an artificial appearance in the lack of sorting of the grains

and the flaky ferruginized surface with the oblique and irregular penetration of the iron oxide, as though a coarse mud had been patted on the leaf, and suggest that it was part of the thatch of a dwelling, held down by mud.

The specimen was highly irregular in shape when collected but was subsequently broken into 3 pieces. It was collected in the coarse lake beds below the humus layer at Tamarindo about 33 ft. above the present level of the Lake and hence comes from the oldest and highest terrace which means that it was deposited toward the maximum expansion of the waters of the ancient Lake and before the oldest terrace scarp was cut during the first stage of recession of the Lake waters.

I know of no means of distinguishing between the leaves of the exclusively American genus *Heliconia* and the generally considered non-American genus *Musa*. If these are *Musa*, which they greatly resemble, they confirm the presence of this genus in America in pre-Columbian time as was clearly shown by the seeds described some years ago from the Tertiary of Colombia.[1]

Geological Sections

I. Section in Railroad cut across the Neck of La Cabrera Peninsula, Eastern end.

Background is high metamorphic ridge, the westward continuation of which forms the rocky axis of the peninsula.

Excavated bench for the Maracay-Valencia auto road.

Humus	4 to 6 ft.	sloping gently southwesterly away from the divide and toward the Lake
Planorbis marl	4 to 5 ft.	
Agglomerate	5 to 10 ft.	
Irregular contact.		
Metamorphics	? ft.	
Track level.		

II. Section at the northern end of same Railroad cut.

Humus with pebbles and marly stringers, with rounded and predominately quartz pebbles at base	5 to 6 ft.
Grayish pebble bed with cobbles	$2\frac{1}{2}$ ft.
Buff or yellowish sand with thin bluish clayey lenses and marl	$3\pm$ ft.
Pebble bed (mostly quartz) becoming sandy and carbonaceous (dark) downward	2 ft.
Calcareous yellowish sand, exposed about	6 ft.
Track level.	

III. Section of easternmost excavation at Tamarindo.

Humus with occasional cobbles (maximum diameters 6 to 8 inches) bits of charcoal and shells	4 ft.
Planorbis marl with scattered pebbles	$\frac{1}{2}$ to 2 ft.
Brownish impure Planorbis marl, with irregular pebbles, much charcoal and bone fragments, thickening toward the Lake	2 to 5 ft.
Pebbles in silty carbonaceous matrix, with bits of broken pottery	1 to 2 ft.
Brownish humus, with occasional pebbles (maximum diameter 3 in.), bits of charcoal. Pebbles are vein quartz and metamorphics, grades downward into grayish impure Planorbis marl, which tends to be in pockets	1 to 3 ft.

[1] Berry, Edward W. *Amer. Jour. Sci.*, **10**, 530–37, 1925.

Angular metamorphics up to diameters of 1 to 2 ft., the larger tend to be at the base with their flat surfaces oriented approximately with the surface of the underlying member .... 4 ft.
Humus, a sandy dark soil and vegetation layer full of Planorbis, exposed .... 2 ft.
All dipping gently toward the present Lake.

IV. Section of the main excavation at Tamarindo.

Dark humus, with many secondary burials .... 2 to 4 ft.
Alternating 2 to 4 inch layers of humus and pebbles with comminuted shells .. 8 ft.
Coarse beds, mostly gravels (both rounded and angular quartz and metamorphics), with large boulders, pockets of ashes and clay full of fish bones, broken pottery especially pipes, human and animal bones, human crania (undeformed) and no secondary burials, turtle plates, the whole with scattered Planorbis shells .... 10 ft.
This horizon rises gently to the southward and on the back face of the excavation, 22 feet distant is 4 or 5 ft. higher, finer grained, and more regularly bedded.
Gray, sandy, somewhat ironstained clay, becoming coarser upward, with pockets and thin lenses of more argillaceous material, with scattered pockets and thin lenses of Planorbis and charcoal .... 13 ft.

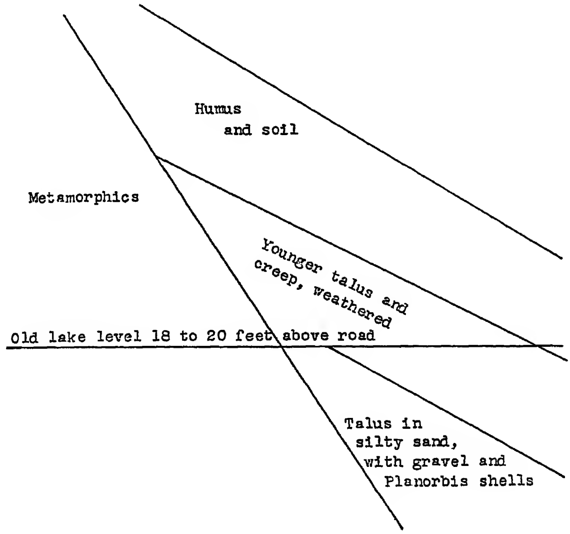
V. Section of test trench 100 feet west of camp.

Continuous humus, formed under emergent conditions .... 2 to 3 ft.
Humus alternating with pockets of yellowish sand and gravel, with sparing Planorbis shells, laid down under alternating submergent and emergent conditions .... 3 ft.
Yellowish sand and gravel, with abundant Planorbis .... 1 to 2 ft.
Human burials are intrusive and extend into the top of this member.
Similar, but with cobbles and boulders, exposed to bottom of trench .... 2 ft.

VI. Section of cut on northeast side of Valencia Road, about 1 km. west of Cabrera RR. station.

Humus, with angular pebbles, and cobbles and broken pottery .... 2 to 4 ft.
Planorbis marl with sparing pebbles, Jacare bones near middle, dipping gently toward the Lake .... exposed 4 to 5 ft.

VII. Section of Road cut, north side of Valencia Road, $3\frac{1}{2}$ miles west of Rio Javilla and 2 miles east of plaza at Guacara.



VIII. Section of banks of Rio Aragua from auto road to its outlet.

Soil and humus, with traces of pottery	6± ft.
Diatomaceous Planorbis marl	10 ft.
Blue clay . . . exposed	4 ft.

These are practically horizontal and appear to form the terrace plain lying east of the Lake and representing the third from the oldest terrace surface on which the so-called mounds of the lake dwellers occur.

Sections of Test Pits (for location see diagram, figs. 1, 3)

Pit 1. 1.8 miles from Highway, 0.7 miles east of Tamarindo.

	Feet	Inches
14. Carbonaceous, pebbly humus	2	
13. Pebbly, oxidized, sandy clay with limonitic streaks		4
12. Comminuted shaly clay layer		4
11. Buff, limonitic clay		3
10. Sandy, gray, slightly oxidized clay	1	2
9. Carbonaceous, black, sandy clay, with charcoal and fragments of pottery (Accession No. 1815 of table)		6
8. Sandy clay with charcoal (Accession No. 1814 of table)		4
7. Diatomaceous earth with Planorbis		6
6. Carbonaceous streak on all 4 walls of pit, charcoal and ashes (Accession No. 1813 of table)		1
5. Diatomaceous earth (Accession No. 1812 of table)	1	
4. More clayey diatomaceous earth (Accession No. 1811 of table)		4
3. Diatomaceous earth (Accession No. 1810 of table)	1	2
2. More carbonaceous layer (Accession No. 1809 of table)		6
1. Light diatomaceous earth (Accession No. 1808 of table)		4

Pit 2. Immediately south of road, Tamarindo.

	Feet	Inches
15. Dark humus with coarse pebbles	1	6
14. Loose coarse clayey gravel	1	
13. Buff clay with gravel		10
12. Same, gray in color and indurated	1	2
11. Loose sand with gravel, very fossiliferous		4
10. Loosely cemented coarse gravel		2
9. Fine, loose, very fossiliferous sand		9
8. Buff, poorly consolidated, ferruginous sandstone		6
7. Fossiliferous, fine, gray sand and gravel		5
6. Soft clay with scattered gravel		8
5. Yellow, iron-cemented, coarse gravel		6
4. Gray, loosely cemented, fossiliferous gravel		3
3. Indurated, iron-cemented gravel	1	4
2. Loose, fine, gray gravel		8
1. Buff, indurated gravel	1	4
Loose gray gravel	1	1
Buff, limonitic gravel	1	2

Bed rock at 12 feet, 8 inches below surface.

Pit 3. Mouth of Tamarindo entry to water level of Lake.

Diatomaceous marl with shells, surface to a depth of 10 feet.
Two feet from the bottom a broken jar about 8 inches in diameter.
Fossils from bottom at 10 ft. are listed in table under Accession No. 1816.

Pit 4. Seven-tenths of a mile east of Saddle in the Peninsula, 50 ft. south of road.

		Feet	Inches
Humus		2	6
Coarse sand with gravel		2	2
Finer, gray sand and gravel		1	
Slightly indurated, soft, gray, sand with large pebbles			4
Diatomaceous earth			5
Compact gray clay with large metamorphic pebbles		1	8
Brownish silt (Accession No. 1817 of table)			4
Compact gray clay			6
Highly fossiliferous gray sand with pockets of Planorbis			5
Gray sand . . . . 2 inches	All layers highly fossiliferous	1	9
Buff " . . . . 4 "			
Gray " . . . . 2 "			
Buff " . . . . 1 "			
Gray " . . . . 5 "			
Buff " . . . . 1 "			
Gray " . . . . 2 "			
Buff " . . . . 4 "			

Pit 5. About 70 feet south of Pit 2.

Soil and humus	1
Sandy gravel	1

Basement rock at about 2 ft. below the surface.

Pit 6. About 70 ft. slightly west of South from Pit 5 starting on next to the younger minor terrace level.

Soil and humus	6–8 inches
Gray sand and gravel, with Planorbis	5 "
Gray clay	8–12 "
Planorbis layer	2–6 "
Diatomaceous earth with streaks of Planorbis	about 2½ ft.
Iron-stained layer with cobbles	15 inches

Basement rock at about 5 feet.

Pit 7. Old well, northwest of camp, Tamarindo.

Humus with irregular cobbles	5 ft.
Medium gravel	14 inches
Clay with Planorbis and irregular boulders	20 in.
Gray, sandy Planorbis clay with charcoal (pipe)	2 ft.
Dark sandy clay with ashes	7 in.
Sandy, fine gravel with much charcoal and human bones	8 in.

Pit 8. Directly south of excavation at Tamarindo.

Humus	16 in.
Sand and gravel, with Planorbis	11 in.
Angular agglomerate with broken pottery, beads, and human bones	5 ft. 6 in.
Coarse gravel with cobbles, much charcoal and pockets of Planorbis	5 in.
Loose gray sand	4 in.
Diatomaceous earth	4 in.
Fine gravel with charcoal and Planorbis	3 in.
Loose, fossiliferous sand	8 in.
Planorbis bed, charcoal and pebbles	5 in.
Gray sand with small pebbles and much charcoal	7 in.
Buff sand	2 in.
Loose gray sand with few Planorbis and many Natica-like shells	9 in.
Brownish diatomaceous earth with Planorbis	1 ft. 2 in.
Gray sandy marl, with charcoal	6 in.

Pit 9. 70 ft. south of Pit 8.

Humus with pebbles	1 ft. 6 in.
Loose gray sand and gravel	9 in.
Black gravel	2 in.
Loose, gray gravel	4 in.
Water-laid black humus and gravel	5 in.
Gray sand, with broken shells	3 in.
Muck (ashes) with large boulders	1 ft. 1 in.
Gravel, with boulders, broken pottery, axe heads, pipes, human and fish bones, deer and Jacare bones	1 ft. 2 in.
Planorbis bed, with ashes and charcoal	1 ft.
Gray diatomaceous Planorbis marl (no trace of humans)	6½ feet
Planorbis bed, almost exclusively shells	1 ft. 1 in.
Loose black gravel (manganese)	3 in.
Loose reddish gravel	5 in.
Gray sandy clay	7 in.
Reddish gravel	6 in.
Compact, barren red clay and gravel	1 ft. 6 in.

Pit 10. About 160 ft. south of Pit 9, surface slightly lower and about 2 feet above the level of the road.

Humus with cobbles	1 ft.
Carbonaceous Planorbis marl with charcoal	2 ft.
Cobble bed with charcoal, Planorbis, broken pottery and bones	1 ft.
Fine gravel	3 in.
Dark sand with pebbles	6 in.
Planorbis marl with small pebbles	1 ft. 8 in.
Dark sandstone with broken shells	1 ft.
White Planorbis marl	6 in.
Soft buff sand with shells	4 in.
Loose, fine black gravel, with broken shells	4 in.
Same, but light in color	5 in.
Black mud (ashes)	9 in.
Gray clay with pebbles and Planorbis	3 ft.
Yellow, indurated, pebbly clay	2 ft.

Invertebrates Collected

All of the forms listed appear to be still existing species. The samples from the test pits show no differences which can be construed as having a chronological significance. For example: the only marked exception to the dead level of uniformity among all of the samples is from Horizon 9 (1815) in Pit 1 from which 11 different species of gastropods were obtained. Only two of these are found singly and sparingly in any other of the samples or lake dredgings, and yet this horizon is one of the more recent horizons in Pit 1, showing conclusively that any significance to be attached to these gastropods is an ecological and not a stratigraphic one.

The most widespread and abundant of the molluscs are *Taphius pronus* (Martens) and *Pomatopurgus parvulis* Guild, and their shells are present in inconceivable numbers in windrows along the beaches, in almost pure beds in the shallower waters, and in comparable stringers

and pockets among the old lake sediments. They are quite as abundant as the Planorbidæ at the classic locality of Steinheim, and doubtless would yield similar results from detailed studies.

Table of Distribution of Invertebrates

A = Abundant, *i.e.*, very common.
C = Common.
R = Rare (Lohman distinguished between Few and Rare, which are treated here as one category).

	Pit 1								Pit 4, Horizon 11	Recent		North Side of Valencia Road 79 8 km.	Marl Overlying Metamorphics LaCabrera R R Cut	Pit 3
										Shallow 1–10′	Deep 100′			
	1808 1	1809 2	1810 3	1811 4	1812 5	1813 6	1814 8	1815 9	1817	1806	1807			Bottom 1816
Drepanotrema anatinum d'Orb.												C		
Unancylus rechnis d'Orb.	C		A	C	C	A	R	R	C	R	C	R		R
Hebetancylus moricaudi d'Orb.	C		A	A	A	A	R		C	R	C	R		
Pomatopurgus parvulis Guild	A	C	A	A	A	A	C	A	A	A	A	C		C
Pomatopurgus valenciæ (Preston)										C	C			
Pomatopurgus ernesti (Martens)										C	C			
Taphius pronus (Martens)	A	A	A	A	A	A	C	A		A	C	A	A	A
Pomacea glauca Linn.							C					R		
Pomacea glauca dubia										R				
Pomacea sp.										R				
Planorbis sp.	R	R	A	C	C	C	R			R	R	C		R
Physa sp.	R	R	C	C	R	C	R		R	R		R		R
Gastropoda (11 different species)	1R						1C	1C 10R		1C	1C			1C
Reptilia														
Jacare sclerops									R					
Fish bones	C		R	R	C	R		A		R	R	R	A	C
Ostracods	A		A	A	A	A	A	C	A	A	C	C	C	A
Seeds								R						C
Chara					R					C	R			A
Diatoms														
Melosira granulala angustissima Müller	R		R		R				R	R	R			C
Cyclotella meneghiniana Kützing											R			
Terpsinoe musica Ehrenberg						R				R	R			R
Fragilaria construens (Ehrb.) Grunow														C
Synedra acus Kützing	R		R	R										
Eunotia monodon Ehrenberg				R										R
Cocconeis placentula Ehrenberg	C	R	C	C	C	C				R	R			C
C. dirupta Gregory var. *flexella* Grunow					R									
Mastogloia smithii Thwaites.	R					R				R				R
Diploneis ovalis oblongella (Naegli) Cleve				R										
Stauroneis anceps Ehrenberg				R										
Anomoeoneis sphærophora (Kützing) Pfitzer	R			R	R									
Navicula sp.														R
Pinnularia dactylus Ehrenberg.														R
Pinnularia nobilis Ehrenberg.				R		R								
Pinnularia sp.	R			R										R
Amphora cf. *A. ovalis* Kützing														R
Cymbella ventricosa Kutzing	C	R	C	C	C	C				R				R
Gomphonema lanceolatum Ehrenberg.			R	C	R	C				R				R
Gomphonema angustatum (Kützing) Ralfs	C	R												
Gomphonema dubravicnesis Pantocsek.	R	R	R	R	C	R				R				R
Epithemia zebra var. *porcellus* (Kutzing) Grunow	C		C	C	C	R				C				R
Rhopalodia gibba (Ehrenberg) Muller	C		R	C	C	C				C				C
Rhopalodia gibberula (Ehrenb.) Muller														R
Nitzschia amphibia Grunow.	R			R		R					R			
Nitzschia sp.				R	R	R								C
Surirella sp.														R

Unknown horizon, mostly surficial, collected by peons in building Indian plaza (Locality 1).

Cricetine rodents
Peccary, *Tayassu torvum*
White-tail deer, *Odocoileus gymnotis*
Crocodilus americanus
Common dog, *Canis familiaris*
Jaguar, *Felis onca*
Swamp deer, *Blastocerus* sp.
Brocket, *Mazama* sp.
Tapir, *Tapirus terrestris*
Armadillo, *Dasypus* sp.
Cinosternum sp. (turtle)
Podocnemis sp. (turtle)
Jacare sclerops
Brazilian cormorant, *Phalacrocorax olivaceus*
Horned screamer, *Anhima cornuta*
Gallinule, *Gallinula chloropus*
Fishes

33 ft. above lake, 10 ft. below surface, west side of entry Tamarindo (Locality 2).

Cricetine rodents
Peccary, *Tayassu torvum*
White-tail deer, *Odocoileus gymnotis*
Common dog, *Canis familiaris*
Rice rat, *Oryzomys* sp.
Swamp deer, *Blastocerus* sp.
Testudo tabulata
Cinosternum sp.
Podocnemis sp.
Nicoria sp.
Jacare sclerops
Piedbilled Grebe, *Podilymbus podiceps*
Brazilian cormorant, *Phalacrocorax olivaceus*
Egret, *Casmerodius alba*
Wood Ibis, *Mycteria americana*
Gallinule, *Gallinula chloropus*
Parrot, *Amazona* sp.
Fishes, very abundant

Humus layer about 1 meter below surface, Kidder excavation (Locality 3).

Cricetine rodents
Capybara, *Hydrochœrus hydrochœrus*
White-tail deer, *Odocoileus gymnotis*
Crocodilis americanus
Spectacled bear, *Tremarctos ornatus*
Common dog, *Canis familiaris*
Crabeating fox, *Cerdocyon thous*
Lizard
Swamp deer, *Blastocerus* sp.
Brocket, *Mazama* sp.
Domestic cattle, *Bos taurus*
Jacare sclerops
Brazilian cormorant, *Phalacrocorax olivaceus*
Horned screamer, *Anhima cornuita*

Mound 53, Locoron, 0.00 to 0.25 meter, Osgood (Locality 4).

Common dog, *Canis familiaris*
Domestic cattle, *Bos taurus*
Jacare sclerops

Mound 53, Tocaron, 0.25 to 0.50 meter, Osgood (Locality 5).

Common dog, *Canis familiaris*
Swamp deer, *Blastocerus* sp.
Jacare sclerops
Horned screamer, *Anhima cornuta*

Mound 53, Tocaron, 0.50 to 0.75 meter, Osgood (Locality 6).

White-tail deer, *Odocoileus gymnotis*
Common dog, *Canis familiaris*
Swamp deer, *Blastocerus* sp.
Jacare sclerops
Brazilian cormorant, *Anhima cornuta*
Graybreasted tree duck, *Dendrocygna autumnalis discolor*
White faced tree duck, *Dendrocygna viduata*
Widgeon, *Mareca americana*
Muscovy duck, *Cairina moschata*
Gallinule, *Gallinula chloropus*

Mound 53, Tocaron, 0.75 to 1.00 meter, Osgood, (Locality 7).

White tail deer, *Odocoileus gymnotis*
Common dog, *Canis familiaris*
Jacare sclerops
Brazilian cormorant, *Phalacrocorax olivaceus*
Muscovy duck, *Cairina moschata*
Limpkon, *Aramus scolopaceus*
Fishes

Locality 8

White tail deer, *Odocoileus gymnotis*

Loose at mound 53, Tocaron (Locality 9).

Bird bones
Jacare sclerops
Fishes

Humus with burials west of camp, La Cabrera (Locality 10).

White-tail deer, *Odocoileus gymnotis*
Bird bones
Common dog, *Canis familiaris*
Jacare sclerops
Piedbilled Grebe, *Podilymbus podiceps*
Gallinule, *Gallinula chloropus*
Macaw, *Ara* sp.

Superficial Indian plaza (Locality 11).

Cricetine rodents
Capybara, *Hydrochœrus hydrochœrus*
Peccary, *Tayassu torvum*
White-tail deer, *Odocoileus gymnotis*
Bird bones
Common dog, *Canis familiaris*
Swamp deer, *Blastocerus* sp.
Domestic cattle, *Bos taurus*
Tapir, *Tapirus terrestris*
Anteater, *Tamandua* sp.
Horse, *Equis caballus*
Jacare sclerops
Fishes

30 ft. above lake, west face Tamarindo (Locality 12).

Capybara, *Hydrochœrus hydrochœrus*
Peccary, *Tayassu torvum*
White tail deer, *Odocoileus gymnotis*
Common dog, *Canis familiaris*
Swamp deer, *Blastocerus* sp.
Jacare sclerops

Cobble layer, 31 ft. above Lake, east face Tamarindo (Locality 13).

Cricetine rodents
Bird bones
Swamp deer, *Blastocerus* sp.
Jacare sclerops
Cocoi Heron, *Ardea cocoi*
Fishes, common

3 ft. below top of coarse beds, Tamarindo (Locality 14).

White-tail deer, *Odocoileus gymnotis*
Bird bones
Jacare sclerops
Redwinged Hawk, *Heterospizias meridionalis*

Bottom of cobble layer, Pit 8 (Locality 16).

Peccary, *Tayassu torvum*
White-tail deer, *Odocoileus gymnotis*
Common dog, *Canis familiaris*
Armadillo, *Dasypus* sp.
Jacare sclerops
Fishes

2 ft. below top of coarse beds, Tamarindo (Locality 17).

Cricetine rodents
White-tail deer, *Odocoileus gymnotis*
Bird bones
Common dog, *Canis familiaris*
Swamp deer, *Bastocerus* sp.
Brocket, *Mazama* sp.
Jacare sclerops
Piedbilled Grebe, *Podilymbus podiceps*
Brazilian cormorant, *Phalacrocorax olivaceus*
Graybreasted tree duck, *Dendrocygna autumnalis discolor*
Fulvous tree duck, *Dendrocygna bicolor*
White faced tree duck, *Dendrocygna viduata*
Black collared Hawk, *Busarellus nigricollis*
Hawk, *Buteo* sp.
Gallinule, *Gallinula chloropus*
Rusty dove, *Leptotila verreauxi*
Macaw, *Ara* sp.
Parrot, *Amazona* sp.
Cayenne Owl, *Rhinoptynx clamator*
Paraquet sp. indet.
Fishes

8 ft. below surface, south face Tamarindo (Locality 18).

White-tail deer, *Odocoileus gymnotis*
Bird bones
Common dog, *Canis familiaris*
Jacare sclerops
Piedbilled Grebe, *Podilymbus podiceps*
Fishes

8 ft. below surface, Tamarindo (Locality 19).

White-tail deer, *Odocoileus gymnotis*
Turtle bones marked by fire
Layers of ashes

3 meters below surface, Tamarindo, Kidder excavation (Locality 20).

Cricetine rodents
White-tail deer, *Odocoileus gymnotis*
Common dog, *Canis familiaris*
Brocket, *Mazama* sp.
Jacare sclerops
Brazilian cormorant, *Phalacrocorax olivaceus*
Wood Ibis, *Mycteria americana*
Curassow, *Crax alberti*
Gallinule, *Gallinula chloropus*
Parrot, *Amazona* sp.
Fishes

Cascabel (Locality 21).

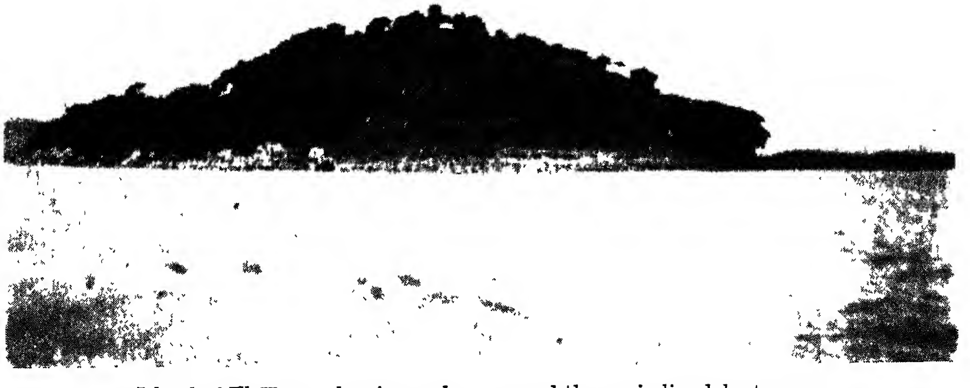
White faced tree duck, *Dendrocygna viduata*

Among other classes of remains, the most abundant, both living and fossil, are Ostracods and Diatoms. Pearse, in his study of the recent lake fauna, enumerated several species of Cladocera and Copepoda, but beyond a few traces of Cladocera I saw no traces of these two orders of Crustacea. The Ostracoda include representatives of the genera *Cryptocandona*, *Dolerocypris*, *Spinocypris*, *Potamocypris*, *Darwinula*, and *Cytheridella*, and according to Blake, *Spinocypris* represents a new genus, and *Cryptocandona*, *Dolerocypris* and *Cytheridella* are represented by new species. These determinations were made from the fossil material, and whether or not they are present in the present fauna of the Lake is not known, although it is highly probable that such is the case.

Among the diatoms the majority from the test pits are found in the Lake in water less than 10 ft. in depth. Fifteen, however, have not been detected in the Recent dredgings. On the whole the diatoms suggest that the old Lake deposits were formed in shallow water, and this is strictly in keeping with their mode of genesis. On the other hand it is worth noting the great abundance of life in the dredgings from the 100 foot depth in the Lake, a much greater depth than at which similar conditions are found in our northern lakes (Juday), which I presume is to be correlated simply with the greater intensity of light in the Tropics.

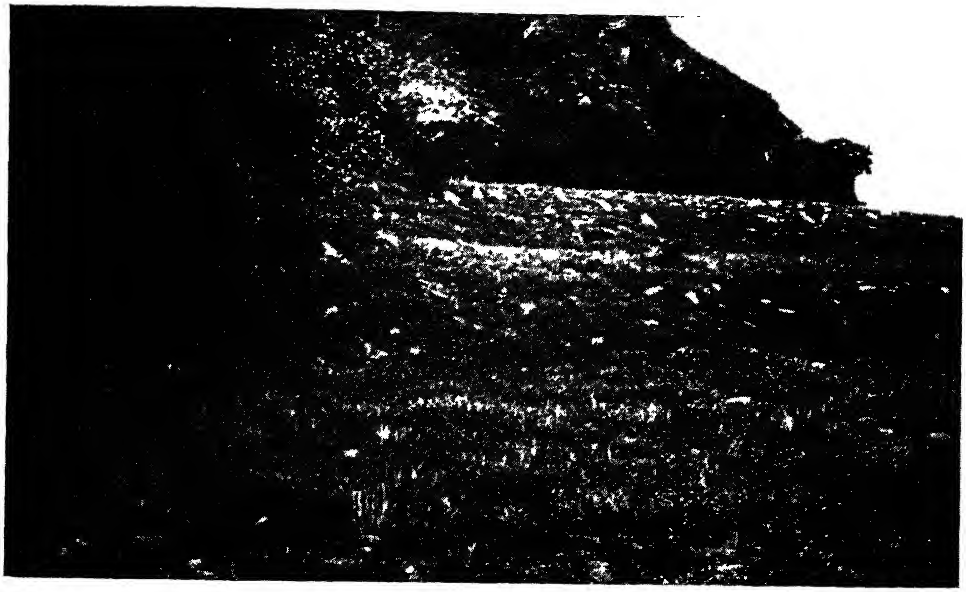
PLATE I

La Cabrera showing the rocky forested core and the arable lake terraces.



Island of El Horno showing rocky core and the encircling lake terraces.

PLATE II



Nearer view of terraces on Island of El Horno.

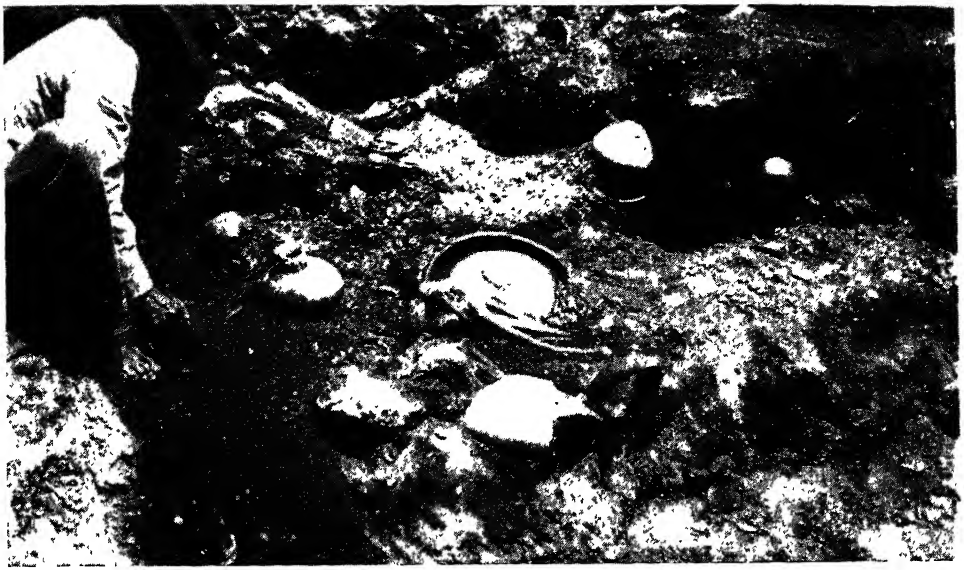
Scarp at base of third level cut in bed rock on El Horno.

PLATE III

South shore of lake at La Florida showing terraces.

Boulders encrusted with calcareous alga *Lithomyxa calcigena* Howe on surface of third level.

PLATE IV



Secondary burials near surface at Tamarindo.

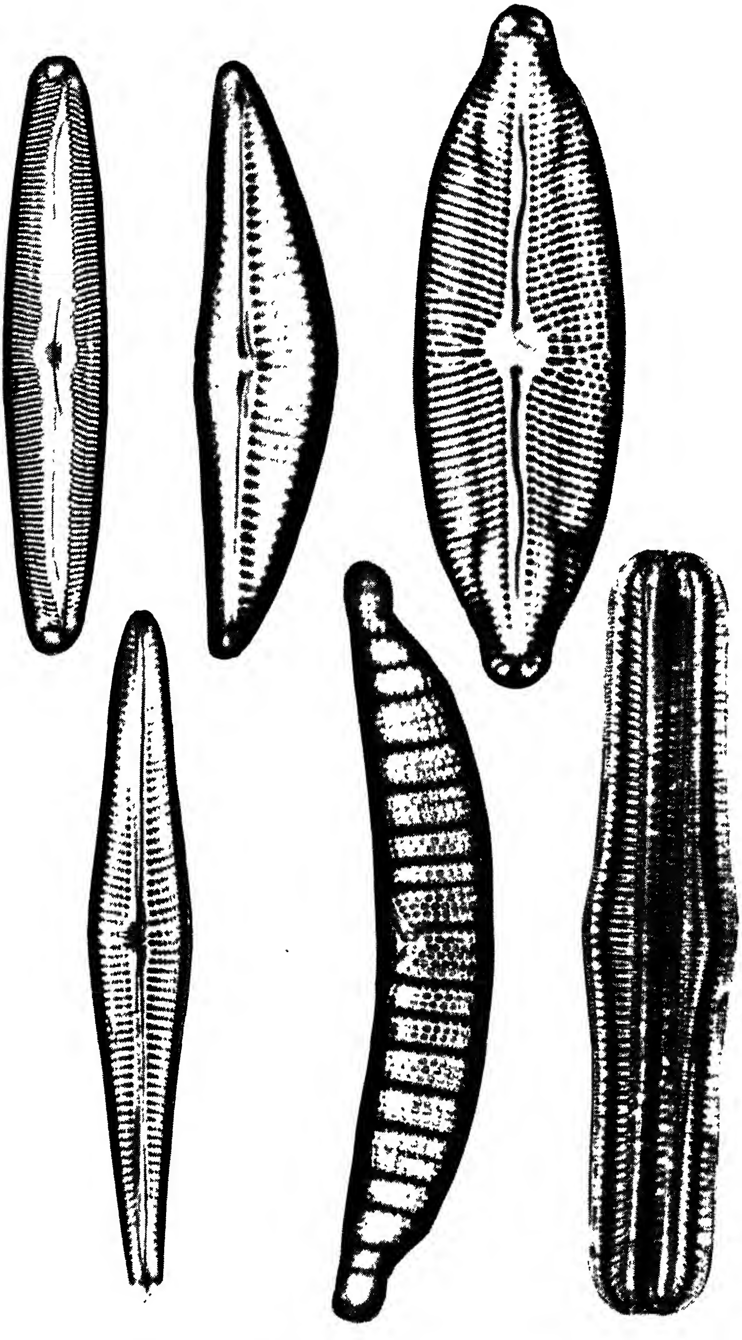
Primary burial at depth of 12 feet at Tamarindo.

PLATE V

Lake beds with human skull 10 ft. below surface.

Coro Indian peon and burial urn.

PLATE VI



1. *Pinnularia viridis* (Nitzsch) Ehrenberg × 680
2. *Cymbella ventricosa* Kützing × 1160
3. *Mastoglœa smithii* Thwaites × 2400
4. *Gomphonema lanceolatum* var. *insignis* Gregory × 1120
5. *Epithemia zebra* var. *porcellus* (Kutzing) Grunow × 1400
6. *Rhopalodia gibba* (Ehrenberg) O. Miller × 920

PLATE VII

THE MASTODONS OF BÅRÅBOEDAER

E. C. CASE

Department of Geology, University of Michigan

(*Read, April 20, 1939*)

ABSTRACT

The great stupa at Båråboedaer near Djaktakarta, in Java, was erected by migrants from India about 800 A.D. The pyramidal structure rises in a series of terraces with high balustrades. The walls and balustrades are elaborately sculptured with scenes from the life of the Bhudda. To permit escape of water from the terraces conduits lead through the balustrades. The outer ends of the conduits terminate in ornamental water spouts in the form of elephant-like heads. These heads have been supposed to represent the mythical Kala Makara, which appears so often and in so many forms in the East Indian mythology. The teeth are, however, typically Mastodon teeth in form and number. This is the sole representation of a dragon form with herbivorous teeth, known to the author. The author speculates upon the origin of this apparent representation of a Mastodon in a structure erected in the 9th century of our era.

It is always a bit dangerous for one to venture beyond his own field in making observations and drawing conclusions therefrom, but the observations described in this paper are such as would not be likely to occur to those interested primarily in architecture or archeology. The visit of a paleobiologist to the great edifice resulted in the discovery of the, to him, most suggestive facts here recorded. Their value remains to be proven.

Those who have investigated agree that the great stupa of Båråboedaer, near Djakjakarta, in Java (Fig. 1, Pl. 1), was erected about A.D. 800 by migrants from India. This gigantic structure is pyramidal in form, is 400 feet square at the base and rises in nine levels or terraces. The combined length of these levels is close to 2 miles and on the walls are 2000 panels illustrating in bas-relief incidents from the life of Buddha. It is, according to Smith in his *History of the Fine Arts in India and Ceylon*, "a sort of book of the life of Buddha according to different accounts."

Each level has a solid balustrade on the outer side which carries a portion of the sculptured panels. In the climate of Java it was necessary to provide for the escape of excess rain water from the different levels. This was done by piercing the balustrades by conduits placed at the corners of the pyramid and at intervals along the sides. The water passed through these from level to level until it reached the ground. The conduits terminate on the outer side in ornamental water

spouts (Fig. 2, Pl. I). Many of the spouts are in the form of elephant-like heads and it is these which are the subject of this paper.

The heads are about half the size of that of an Indian elephant, the trunk is raised, carrying some object which is perhaps a formalized bunch of fruit or grain, perhaps a jeweled tassel. The tusks are obvious and the mouth widely opened to permit the efflux of the water. The widely opened mouth reveals the carefully shaped cheek teeth which are typically those of the extinct mastodon. This is particularly apparent in the upper jaws, though the lower teeth are more mastodon-like than appears in the accompanying photographs (Figs. 1 and 2, Pl. II). There are four cheek teeth in each jaw, each tooth with a cingulum and 2 or 3 cross ridges which would have engaged with the ridges of the opposing tooth in occlusion. These characters are notably different from those of the dentition of the Elephant where there is but a single functional cheek tooth in each jaw, four in all, no cingulum, and the top flat and crossed by many bands of enamel. The Mastodon was an herbivorous, browsing animal the jaws of which moved up and down, the Elephant feeds upon hard grains and grasses which it grinds between the teeth that move forward and back against each other. Even the least biologically minded people could not fail to note the marked difference.

The migrants from India knew the elephant and it is hardly reasonable to suppose that they would have carved so different a tooth by error. If the choice lies between the two forms, Mastodon or Elephant, it seems far more reasonable to assume that the sculptors were attempting to represent an archetypal form of proboscidean suggested by the discovery of a fossil Mastodon. Such a skull was perhaps preserved in some temple and its characters persisted in a stylized religious art. Remains of the Mastodon have been found both in India and Java. They are of the Tetralophodont type in which the cross-ridges on the teeth are more numerous and more tuberculate than in the American Mastodon, but very possibly the suggested models of the teeth in the heads at Båråboedaer.

The Mastodon is supposed to have become extinct long before the time of the building of the stupa of Båråboedaer, but there are authentic accounts, as from South America, of its remains being found with artifacts and other evidence of the struggle with human enemies which ended in a gargantuan feast. The possibility that the Mastodon survived in some remote locality until the ancestors of the builders of Båråboedaer saw it and preserved an account of it in legend is not too fantastic to be considered.

There is the possibility that the teeth in the elephant-like heads were copied from those of the East Indian tapir which are also cross-ridged, but as the tapir has no tusks, sharp, erect ears, and only a prehensile upper lip, it seems a rather forced explanation, even considering the vivid imagination of the East Indian artists unhampered by respect for morphological relevancy.

An alternative explanation, which seems more acceptable to the Dutch scientists of Java whom the author was enabled to consult indirectly through the kind aid of the late Doctor Charles C. Blackshear, an American long resident in Djakjakarta and well acquainted with Indian and Javan antiquities, is that these ornamental water spouts are but variants of the abundant representations of the legendary Makara. The Makara was a rather Protean monster which appears frequently in Indian art and wherever Indian art and mythology spread. It is described by Coomaswaramy [1] as follows: "The Makara has the trunk of an elephant, the feet of a lion, the ears of a pig, the body of a fish living in the water, the teeth turned outward, eyes like Hanumans, and a splendid tail." Figures of the Makara by the same author [2] show a creature apparently aquatic with a long trunk, erect canine teeth above and below, cheek teeth numerous and bluntly pointed, short front limbs and digits, no hind limbs, tail ending in convolutions which seem to shade into conventionalized waves of water, and large elephant-like ears. Other writers and figures give the Makara the body of an antelope, the body and tail of a bird, the head of a dolphin, etc. Evidently the Makara was figured much as the artists' fancy dictated but it seems commonly to have retained the elephant-like proboscis, the tusks and the fish-like tail, as determining attributes. The archeologist would most naturally consider the ornamental water spouts at Bårâboedaer as a variant of the Makara retaining the trunk and tusks but with the tail absent as only the head was used. In opposition to this view one thing appeals strongly to the author. In none of the figures or descriptions of the Makara that he has been able to examine are the cheek teeth other than the conical fangs of a carnivorous creature. There is a picture in Coomaswaramy's *History of Indian and Indonesian Art*, Pl. LXXVII, Fig. 249, of a heavy bodied animal like an elephant, but with the heel high on the leg as in ungulates, a long trunk or upper jaw and with the cheek teeth of the upper jaw bilobate. This is from Bijapur in India, dated about A.D. 1100. Rea also refers [3]

[1] Coomaswaramy, A. K., *Mediaeval Sinhalese Art*, p. 84. With numerous illustrations of the Makara.

[2] Coomaswaramy, A. K., *History of Indian and Indonesian Art*, 1927. Notably Pl. XLVII, Fig. 177.

[3] Rea, S., *Indian Buddhistic Antiquities*, Pl. XI.

to this or a similar representation with a long upper lip or trunk, no tusk and bifurcate teeth. These are the sole examples, except those at Bårâboedaer, known to the author, in his limited examination of the subject, where dragon-like forms are represented in India or elsewhere with other than fang-like, carnivorous teeth. The common conception of such things seems to have been of a fierce and predatory creature. The sculptured heads at Bårâboedaer are unique in representing an herbivorous animal and peculiar in that they represent an elephant-like animal very different from the Asiatic elephant with which the builders of the stupa must have been familiar Such a fundamental departure from a long established conception and fixed convention of artistic representation is easier to explain as due to an objective suggestion such as a sight of a mastodon skull, than by the vagaries of imagination in a single artist.

In closing, it is interesting to note that the architects of the stupa used the Makara exactly as the Mayas and Aztecs used the feathered serpent, the kukulcan or quetzalcoatl. It is traced on the borders of the niches sheltering statues of Buddha, the head below and the long tail rising on the sides to entwice in the lintel with that of the one on the opposite side. The stairs have heads at the lower end, as newel posts, and the long body rising as the balustrade.

FIG. 1. Båråboedær, near Djakjakarta in Java. This and the accompanying figures are taken from Krom, N. J., Bechrijving van Båråbadaer samengesteld door, N. J. Krom en T. van Erp. 's Gravenhage 1920–31.

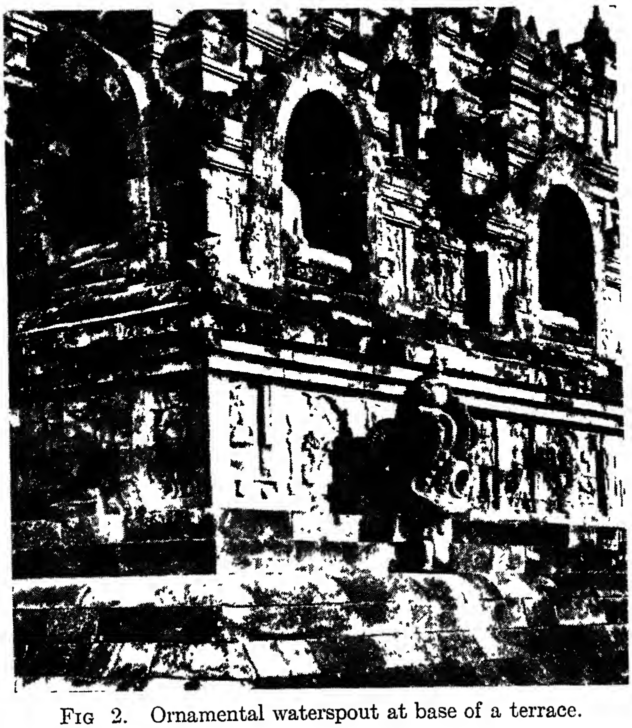


FIG 2. Ornamental waterspout at base of a terrace.

Fig. 1. Side view of an ornamental waterspout showing tusks and mastodon-like teeth.

Fig. 2. Front view of another ornamental waterspout showing the tusks and mastodon-like teeth. In this case there is a smaller dog-like figure in the open mouth.

NOTES ON SHAKESPEARE

WILLIAM LYON PHELPS

Professor Emeritus of English Literature, Yale University

(*Read April 21, 1939*)

THESE are merely scattered notes and observations. Although Shakespeare was married to Anne Hathaway when he was eighteen years old and she was eight years older than he and their first child was born about six months after the marriage, that is almost all we know about their relations. In a certain sense these facts speak for themselves, but there is no evidence of any quarrel or even coolness between them. She survived him and in his will he left her the second best bed. Many people have taken that bequest to be an insult when it is far more likely to have been a token of the most intimate affection. Even until fifty years ago, the spare room or guest room in the house was usually the best and it probably contained the best bed. The second best bed, therefore, would have been a sign of the most intimate affectionate relationship. This interpretation is strengthened by the will of Boccaccio who in his will left parts of beds to family relatives.

One of the most astonishing things about Shakespeare's miraculous career is its brevity, extending from about the year 1593 to 1610 or 1611, a little less than twenty years. One of the greatest enigmas of his life is his ceasing to write when he was about 46 years old, leaving London and returning to Stratford. A good many conjectures have been made about this. The least probable of all is, I think, the one suggested by Lytton Strachey, who said that it was probable that Shakespeare had lost interest in life. That seems to me incredible. Among so many conjectures I should like to put one of my own. The Sonnets were published for the first time in 1609. Although these are full of allusions that nobody understands today, they must have been perfectly clear to all the people who read them in London at the time of their publication, and whatever view we may take of the Sonnets as to whether they represent actual events or whether they are purely imaginary, their publication must have been rather awkward for the author. Take, for example, the one beginning, "No longer mourn for me when I am dead." In addition to that and many other allusions there is one sonnet that is wholly obscene and others that allude to

various rather scandalous affairs. Now it is just possible that when this book was published, Shakespeare decided to leave town. Stratford is a little over one hundred miles from London and in those days that meant a pretty effectual seclusion from the life of the city.

One of the greatest examples of irony of history is shown in the repeated statements in the Sonnets that Shakespeare is giving immortality to the *name* of his friend. In those days, of course, everyone knew who that friend was and now nobody knows. Therefore, Shakespeare, conferring immortality on himself by the genius contained in these poems, has succeeded in keeping his friend's name in complete oblivion.

In general the two opposite interpretations of the Sonnets may be shown in the opinions of two of the foremost living Shakespeare scholars in the world, Professor Kittredge of Harvard and Professor Tucker Brooke of Yale. Mr. Kittredge believes that we can learn little of the facts of Shakespeare's life from the Sonnets—that they are purely the work of imagination. Mr. Brooke expresses in his recent edition of the Sonnets that they are filled with definite allusions to actual happenings. There you have two irreconcilable interpretations from two scholars who are respected by all students.

Although I am not sure that actual events are described in the Sonnets, there are certain interesting deductions that we can make. For example, Shakespeare must have suffered terribly from the English climate. He speaks so often of the tragic shortness of summer, of the harsh winds even in May, of the frequent rains even when the morning was promising. I think he must have been unusually sensitive to cold and to clouds and to storms. Thirty-three sonnets have reference to the weather. A little more than eighty years after his death another famous dramatist, William Congreve, in replying to the Rev. Jeremy Collier's attack on the inordinate profaneness and immorality of the theatre, said that the English climate was so abominable, so cold, so chill, so dark, and so depressing, that the writers for the theatre were justified in immoralities in the endeavour to get some comic relief against the depression caused by the climate. This is the only instance I know of in literature where the climate has been held as an excuse for published immoralities. One would think that it would be more natural if it caused profaneness rather than immorality.

There is only one department of literature in which Shakespeare was second-rate, and that is in narrative poems. In 1593 he published *Venus and Adonis* and in 1594 *Lucrece*. He himself seemed extremely proud of these as he prepared them for publication and dedicated them

to the Earl of Southampton, but if he had never published anything else he would be now entirely forgotten. Compared with the Sonnets and with the best passages of the plays they are commonplace.

On Saturday, 25 April 1778, Johnson and Boswell had a discussion about the meaning of the word pamphlet, as follows:

> I happened, I know not how, to say that a pamphlet meant a prose piece. Johnson. "No, Sir. A few sheets of poetry unbound are a pamphlet, as much as a few sheets of prose." Musgrave. "A pamphlet may be understood to mean a poetical piece in Westminster-Hall, that is, in formal language; but in common language it is understood to mean prose." Johnson (and here was one of the many instances of his knowing clearly and telling exactly how a thing is). "A pamphlet is understood in common language to mean prose, only from this, that there is so much more prose written than poetry; as when we say a *book*, prose is understood for the same reason, though a book may as well be in poetry as in prose. We understand what is most general, and we name what is less frequent."

Johnson could have absolutely clinched his argument had he been familiar with Shakespeare's narrative poem, *Lucrece*, because the dedication of this poem, signed by William Shakespeare, begins as follows:

> The love I dedicate to your capital lordship is without end, whereas this pamphlet, without beginning, is but a superfluous moiety.

But although Johnson in his edition of Shakespeare wrote one of the most famous commentaries in history, it is probable that he had never read Shakespeare's Sonnets or the two narrative poems but was familiar only with the plays.

In the great declamation by the professional player in the presence of Hamlet (the lines beginning "the rugged Pyrrhus") it will be remembered that the player is unable to continue because of his emotion. His voice breaks and his eyes are filled with tears. In Hamlet's soliloquy which follows he condemns himself for not being more distressed over the death of his father when the player is overcome with grief at the death of the purely imaginary Hecuba.

Personally I don't think the player was overcome because of any sorrow about Hecuba. What put a lump in his throat and tears in his eyes were the majesty and splendour of the lines. There are passages in literature which none of us can read aloud without being overcome by emotion. I remember when John Masefield was asked to read aloud his poem, *August 1914*, he broke down in the middle of it and said he would have to read something else. It is, then, the poetry and not the death of the imaginary queen that broke the actor's self-control.

Whatever social position Shakespeare may have had in Stratford

or during the earliest years that he spent in London, it is certain to my mind that he very soon had for his intimate friends social aristocrats and that he made visits in splendid homes. For example, in the famous Song on Winter in *Love's Labour's Lost,* the tone is patrician; the affectionate intimacy in the attitude toward servants is the true patrician attitude. The man who wrote this was used to being waited on by servants in the hall and in the kitchen.

When Isicles hang by the wall,
And Dicke the Sphepheard blowes his naile;
And Tom beares Logges into the hall,
And Milke comes frozen home in paile:
When blood is nipt, and waies be fowle
Then nightly sings the staring Owle
Tu-whit to-who.
  A merrie note,
  While greasie Jone doth keele the pot.

When all aloud the winde doth blow,
And coffing drownes the Parsons saw:
And birds sit brooding in the snow,
And Marrians nose lookes red and raw:
When roasted Crabs hisse in the bowle,
Then nightly sings the staring Owle,
Tu-whit to-who:
  A merrie note,
While greasie Jone doth keele the pot.

The famous death scene of Sir John Falstaff in *King Henry V* is additionally famous by containing what is probably the most felicitous emendation ever made, that by Theobald in the eighteenth century. In the following passage the line is "For his Nose was as sharpe as a Pen, and a Table of greene fields."

Nay sure, hee's not in Hell: hee's in *Arthurs* Bosome, if ever man went to *Arthurs* Bosome: a made a finer end, and went away and it had beene any Christome Chile: a parted ev'n just between Twelve and One, ev'n at the turning o' th' Tyde: for after I saw him fumble with the Sheets, and play with Flowers, and smile upon his fingers end, I knew there was but one way: for his Nose was as sharpe as a Pen, and a Table of greene fields. How now Sir *John* (quoth I?) what man? be a good cheare: so a cryed out, God, God, God, three or foure times: now I, to comfort him, bid him a should not thinke of God; I hop'd there was no neede to trouble himselfe with any such thoughts yet: so a bad me lay more Clothes on his feet: I put my hand into the Bed, and felt them, and they were as cold as any stone: then I felt to his knees, and so up-peer'd, and upward, and all was as cold as any stone.

Every commentator has assumed that "a Table of greene fields" is nonsense. Theobald suggested "a' babbled of greene fields." Whether Shakespeare wrote this or not, the line is fully worthy of his genius. Some forty years ago it was an elocutionist, Mr. Locke Richardson, who for the first time suggested a new explanation for this passage. Ordinarily we do not expect many original ideas from elocutionists, but Mr. Richardson made a most valuable contribution to Shakespeare when he said that the phrase "a' babbled of greene fields" meant that Falstaff in his dying delirium was trying to repeat the Twenty-third Psalm—"He maketh me to lie down in green pastures." Now if this be so, I should like to make a still further suggestion that the word "Table" may not be nonsense at all, for the Twenty-third Psalm mentions the table—"He prepareth a table before me in the presence of my enemies." Why, therefore, may we not adopt the suggestion of Mr. Richardson that he was repeating the Psalm and yet preserve the text, that is "a Table of greene fields," thus giving additional evidence that it was the Twenty-third Psalm?

Shakespeare signed his will on the twenty-fifth of March, 1616, and died less than a month later. It is possible at least that he knew he was fatally ill, for the story that he died of a drinking bout did not appear until fifty years after his death and does not deserve credence.

We know in Hamlet's advice to the players when he spoke of Herod that he was alluding not to the Herod of the Bible but to the Herod of the mystery plays which he had seen. Another interesting point here is that Bacon speaks of a jesting Pilate. There is not the slightest evidence in the Bible that Pilate was jesting, but in some mystery plays he was. The study of certain words in Shakespeare is interesting. For example, the exclamation "Fie"; that word today has lost all its power. To say "Fie, fie," is about the same as saying playfully, "Naughty, naughty," but "Fie" was a tremendous and terrible word in Shakespeare, about the equivalent of the French "Pouah!" For example, Hamlet in the first soliloquy in the play, speaking of the world, says, "Fie on it, ah, fie, 'tis an unweeded garden." The word "Fie" to Elizabethan ears must have been an expression a hundred times stronger than it became in later years.

Finally, as is well known, Shakespeare attained fame and popularity immediately. His contemporaries regarded him as the most famous and popular playwright of the day, and some years after his death Ben Jonson said that he had surpassed all the writers of Greece and Rome, and Jonson was a fanatical admirer of the ancient classics. But what particularly impressed his contemporaries was his ease and fluency, not

having to wait for inspiration, and it was even a fact that he did not have to revise or correct his lines. As Ben Jonson said, he never blotted a line, although Jonson thought he ought to have done so.

When young Milton wrote the commendatory verses to the second Shakespeare folio in 1632 (Milton was twenty-two or twenty-three), he used this expression:

> For whilst to the shame of slow-endeavouring art
> Thy easy numbers flow.

Ben Jonson was living when these lines were published. And when young Milton wrote the poem, *L'Allegro*, he made a comparison between Shakespeare and Jonson that was understood by everyone.

> Then to the well-trod stage anon
> If Jonson's learned sock be on,
> Or sweetest Shakespeare, Fancy's child,
> Warble his native wood-notes wild.

That line was worth a whole library of criticism, and for a century after Milton's death it was constantly quoted.

Among the great English poets Chaucer and Tennyson were alike in constantly revising, correcting, and improving their work, while the great impromptus were Shakespeare and Browning who wrote hot from the brain.

For the first time in the Western hemisphere, the English actor, Maurice Evans, last October presented the complete version of *Hamlet* without any cuts, and it was more exciting, more interesting than any previous production that we had ever heard. There was one moment that particularly interested me during the famous soliloquy beginning, "To be or not to be." Ophelia is on the stage, although Hamlet does not see her. There is a little prie-dieu in one corner of the stage where Ophelia is praying with her back to the audience. Then at the end of the soliloquy Hamlet sees her and says, "Nymph, in thine orisons, be all my sins remembered." I was very glad that Mr. Evans made this a serious request and did not speak it smilingly or playfully. Hamlet is in danger of committing suicide and he asks Ophelia to pray for him. I have always believed that this was the true interpretation of the passage, but I had never seen it so represented on the stage until last October.

In the year 1911 I had a long conversation with the German dramatist, Gerhart Hauptmann, in Berlin. Among other questions I asked him, "Do you express your own opinions in your plays?" He replied, "No, I never express my opinions through any of my characters; but

any person who read half a dozen of my plays ought to know the kind of fellow that I am." The personality of Shakespeare is more difficult to understand than the personality of Hauptmann; but to a certain extent it is possible to get some notion of his character and personality from his plays. Caroline Spurgeon in her recent book, *Shakespeare's Imagery,* took the original and fruitful method of discovering from his metaphors and similes the kind of things that he was interested in, like games and flowers and various avocations. I have in mind something different from that, remembering that while most of Shakespeare's contemporaries died young, probably owing to the reckless manner in which they lived, Shakespeare himself advanced steadily during this same time, not only to increasing fame but to financial security, so that he was able to buy the best house and grounds in Stratford and to write the word "gentleman" after his name. I suggest that he who saw men ruined by excesses all around him, excessive dissipation, excessive bigotry, loved above everything else moderation. I think his ideal man was one characterised above everything else by self-control and what we call "character." It seems to me that I hear Shakespeare's own voice in Hamlet's address to the players where it is clear that he hated bombast and excesses in declamation and preferred intelligent moderation, and we can go farther than that in the same play. Hamlet loved and admired and respected Horatio more than any other man of his acquaintance. This was a college friendship. They were both undergraduates; and Horatio's presence at court gave Hamlet his only comparatively happy moments. Now Horatio was not a nobleman and not a courtier, held no office and had no title, and yet it will be observed that he was trusted by every person who had anything to do with him. He was trusted not only by Hamlet but by the King and Queen.

It is interesting to observe that immediately after the prose speech of Hamlet to the players he addresses Horatio in the following terms:

> Since my dear soul was mistress of her choice
> And could of men distinguish, her election
> Hath seal'd thee for herself. For thou hast been
> As one, in suff'ring all, that suffers nothing;
> A man that Fortune's buffets and rewards
> Hast ta'en with equal thanks; and blest are those
> Whose blood and judgment are so well commingled
> That they are not a pipe for Fortune's finger
> To sound what stop she please. Give me that man
> That is not passion's slave, and I will wear him
> In my heart's core, ay, in my heart of heart,
> As I do thee.

A SOLDIER'S MARRIAGE CERTIFICATE IN DIPLOMA FORM

HENRY A. SANDERS

Professor of Latin, University of Michigan

(*Read, April 21, 1939*)

Abstract

This document, a papyrus fragment from the University of Michigan Excavation at Karanis, is dated in the second century A.D. both by the writing and by the other documents found with it. In form and arrangement of text and signatures it is a parallel to Michigan Inventory 508–2217 recently shown to be from a papyrus diploma. A second example aids in establishing a form of document long assumed from the meaning of the word *diploma*.

The marriage contract of a soldier at this early date has great importance for this much discussed problem, since it indicates that actual marriages occurred and at such late dates in the lives of the contracting parties, that the chief object may well have been the attainment of Roman citizenship by the wife and children upon the honorable discharge of the soldier.

The fragmentary signatures of seven witnesses, seemingly Roman citizens, but written in Greek, appear on the back of the document.

THE papyrus, Michigan Inventory 4703, was found in the University excavation at Karanis in 1926. The excavation number is 26-B17 F-A. In the same room fifty-three other papyrus fragments or groups of fragments were found and nine more in another room of the same house. All on the basis of the writing have been assigned to the second or early third century. One falls in the reign of Trajan, two of Hadrian, seven of Antoninus Pius and ten of Marcus Aurelius. This fragment is on an excellent piece of papyrus with the first two lines rather damaged and terminates apparently at about the middle of the document in a straight line break. In form it is an exact duplicate of the papyrus diploma Mich. Inv. 508–2217 published in *Transactions of the American Philological Association,* LXIX, 104 ff. The contract is written across the fibers in the same manner as in Mich. Inv. 508–2217 and so the signatures of the witnesses on the other side, placed at right angles to the text, are likewise written across the fibers. The finish of the papyrus is so excellent that I cannot distinguish recto from verso, and so thin that the fibers running in both directions can be seen on both sides. The size of the fragment is 26 by 9 cm. (10 by 3½ inches). The lines of writing are from 21 to 22 cm. long. The margins are well preserved, the left hand one being 3 cm. wide and the right hand one varying from 1 to 2 cm. It seems probable therefore that the top margin was at least 3 cm. wide originally. This estimate is supported

by the condition of the Greek signatures on the back of the document. No part of any of the prænomina is preserved, and in all except two of the signatures part of the nomen is also missing. The signatures are in seven different hands and the writing differs somewhat in size. From the analogy of Mich. Inv. 508–2217 we assume that the prænomina were written in full. As the common prænomina vary in length from 5 to 8 letters, it is likely that the longer ones required enough space beyond that needed for the shorter and more common ones, so that in two cases the nomen came so far to the right that it is almost completely preserved. In the other five signatures we may assume two or three letters lost from the nomen. This means that seven or more letters are missing at the beginning of each signature, for which a space of at least 3 cm. would be needed in this type of writing. As the seals which would naturally precede are also missing, another centimeter must be added to the width, making 13 cm. the width of the original document as folded, and 26 by 26 cm. the size of the papyrus sheet used. The clean, straight break at the end of the fragment came from the fold of the diploma at that point, as has been noted in the discussion of Mich. Inv. 508–2217. There the signatures were written each in a single line, except where there was not room. In this diploma each signature takes two lines, the break being made after the tribal designation, which in three cases seems to have a termination or abbreviation stroke. In the first two signatures there was not room to write the cognomen on the same line as the rest of the name, and the other witnesses may have used a second line for the cognomen in imitation of the first two signers, but it seems more probable that all were influenced by the desire to have a break after the tribal abbreviation.[1] Two of the witnesses ended their names with a stroke reaching to the edge of the papyrus. Similar but shorter strokes end the other signatures. One feels here as in Mich. Inv. 508–2217 that the signatures of the witnesses were expected to cover the whole of this side of the diploma.

The damaged condition of the beginning of the document came from the intentional or accidental tearing off of the row of seals and threads that served to bind together the top and bottom of the papyrus sheet after it was folded. Mich. Inv. 508–2217 shows similar damage at the beginning. This agreement in form suggests that other Latin double documents should be reexamined.

The writing of the document is a beautiful maiuscule cursive ap-

[1] In Mich. Inv. 508–2217 there is a similar two-line arrangement of the one signature which contains the tribal abbreviation, though there the cognomen, which is entirely lost, may have been too long for the space at the end of the line.

parently of the early second century. The forms of several of the letters, as *c*, *d*, and *s*, show wax tablet influence. The scribe probably was accustomed to write on both materials. A few letters, as *e* and *m*, favor an earlier date, but *g*, *p*, *q*, and *t* seem of the first half of the second century, a date supported by the whole appearance of the writing, and by the dated documents found in the same house. The Greek signatures on the back, though badly written and showing signs of Latin influence, are not inconsistent with this date.

There can be little doubt that this is a marriage contract, though the actual word meaning "betrothed" or "married" has to be supplied from very meager traces.[1] Yet the contract is in some respects more like the Greek type than the Roman form found in Mich. Inv. 508–2217 and *Papiri della Societa Italiana*, 730, for it omits the essential phrase *liberorum procreandorum causa* as well as reference to the Julian law. On the other hand, the woman, accompanied by an attendant and in the presence of her guardian, gives herself in marriage. In Greek documents a parallel occurs in *Papyri zu Giessen* 2.8, but Mitteis, *Papyruskunde, Juristischer Teil*, I, 216, suggests the Macedonian origin of the woman as a possible explanation. Spiegelberg, *Papyrus Libby*, p. 1 ff., notes the same usage in two early Demotic marriage contracts, P. Libbey 1 and Berlin Papyri, Inventory 3078. An Egyptian origin of the practice seems the more likely for these three examples, but hardly for the Latin contract under discussion. Here it seems the influence of Roman law must have been paramount. Naturally examples of similar marriage contracts can not be cited but we find abundant examples of Roman women accepting inheritances and making contracts, purchases, and sales either with the aid of a guardian or without. Examples are P. S. I. 1027 (Herennia Helene, with guardian present, accepts an inheritance), Bruns, *Fontes Juris Romani*, 123 (a woman attended by the witnesses presents the will of her husband to the magistrates), Bruns, *op. cit.*, 124 (woman with aid of a *procurator* accepts inheritance), and Bruns, *op. cit.*, 133 (a woman without presence of guardian being noted buys half of a house).

Furthermore, in this contract probably the wife and certainly her attendant and guardian are identified by elaborate descriptions including age and physical characteristics. This is a Greek method found in many documents from Egypt.

Still more puzzling from a legal point of view is the fact that the husband and wife had been previously married and had two children,

[1] Professor H. J. Wolff has pointed out the possibility under Roman Law of a dowry contract between husband and wife after marriage, but we have no information about the form of such contracts.

one fourteen years of age and the other ten, to which we must add that the husband is a soldier of the fleet still in active service. No mention is made of a divorce having occurred, though in the case of captivity or disappearance of a soldier the wife could claim a divorce and remarry after five years (*Digest*, XXIV, 2, 1; 6). Roby, *Roman Private Law*, I, 133, assumes that upon the return of a captive after five years' absence, the marriage did not naturally continue but a new consent was necessary.

It is not, however, necessary to assume that the former marriage took place before the soldier enlisted, or that a divorce had occurred, for one might remarry with a written marriage contract the wife with whom he had been living in an unwritten marriage (γάμος ἄγραφος); cf. *Berline Griechische Urkunden* 183.

More important is the fact that this is the marriage of a soldier in active service in the fleet and that the document is not later than the second century A.D. It thus falls in the same period as Papyrus Cattaoui from which Mitteis, p. 281 ff., finds support for the complete prohibition of soldier marriages. Lesquier, *L'Armée Romaine d'Egypte* (1918), p. 262 ff., using the same six legal decisions of P. Cattaoui, holds that *conubium* was impossible for the soldier at least up to the middle of the second century A.D. but he points out that this did not prevent marriages of soldiers nor even the attempts to secure therefrom the results of legal marriages. That marriages of this kind occurred among soldiers generally is shown by the numerous military diplomas, which grant citizenship also to the wife and children of the soldier who receives an honorable discharge. The attainment of citizenship is presumably the chief material gain which the common law wife could hope to get from such a marriage, and it seems fairly clear that the government accepted any form of marriage or mere co-habitation, if it lasted up to the honorable discharge of the soldier. The cases cited by Lesquier, which came before the prefect for decision, with one exception dealt with soldiers who had not received an honorable discharge. Marriage contracts are not cited, but the former wives seek the return of loans or deposits, which are held by the prefect to be disguised dowries. The marriage of Demetria and Valerius Gemellus has nothing similar to these cases. The marriage here repeated for whatever cause is not a disguised attempt to protect the dowry of the woman, though it will protect it after the discharge of the soldier, but, if we may judge from the age of the children mentioned, looks toward the not too distant honorable discharge, which will reward the wife and children herein acknowledged, as well as the soldier himself. For this purpose one form

of marriage is as good as another, provided that it contains the acknowledgement by the husband and is duly witnessed and sealed. The regular form of Roman marriage contract could not be used, since the lex Julia had no mention of soldiers' marriages. Under the circumstances any form of marriage recognized in the locality was a satisfactory substitute. Yet the contract was written in Latin and sealed by seven witnesses as in regular Roman marriage contracts, since its object was to secure Roman citizenship for the wife and children as well as dowry rights. If later a legal question arose a Latin document would probably be more effective before a Roman magistrate.

Text

[. . . . . Lu]c̣ị filia Demetrịa [. . . . .] c̣um coṃi[te] Tịcol
[. . . . . . .] ạnn(is) XXXVIIII corpore fusco fa[c]ịẹ [d]ẹḍ[u]cta naso recto ḷentịc̣o maḷo
p̣[ac]ṭa ẹ[st] ṭutore auctọre Glaucio Pạn[no]niani c̣onsọcolẹḍẹṭạṛ ann(is) XXXXVIII cor
pore fusco facie deducta naso recto subcalvo cicatrice supra super
[c]ilium sinistrum. C. Valerio Gemello mil. classis. Aug. Alexandrinae
Ḷịbyrni Dracontis cui ante nupta erat ex quọ matrimonio fịlios pro
creaverunt. Iustum ann XIIII. Gemellum ann X. eique dotis suae
nomine dixit deditque in aestimio vestis et in ṇumerato praesens
ọc̣ṭ[ingent]ạṣ ḍụạṣ ḍṛạc̣ṃas quam dotem dixit se is Valerị[u]s Ge-
[mellus accepisse. . . .]

Translation

Demetria, daughter of Lucius . . . with her attendant Ticol[louthos] thirty-nine years old, of dark complexion, thin face, straight nose, and freckly cheek and with Glaucius . . . son of Pannonianus, thirty-eight years old, of dark complexion, long face, straight nose, rather bald and with a scar above the left eyebrow, acting as guardian, has betrothed herself to Gaius Valerius Gemellus, a soldier of the Augustan Alexandrian fleet and of the Liburnian galley Dragon, to whom she had been previously married, from which marriage there were born sons, Justus fourteen years old and Gemellus ten years old, and she has promised and has given to him as her dowry clothing by valuation and cash in counted coin eight hundred and two drachmas, which dowry the [said] Valerius Gemellus has acknowledged that he [has received].

1 The document starts with the name of the woman in due Latin form, as, for example, [*Sempronia Lu*]*ci filia Demetria*. In Greek and Demotic marriage contracts, as noted above, a woman is at times recorded as giving herself in marriage; but this seems more

in accord with Roman custom, though exact parallels are lacking. The addition, moreover, that it is done under the guardianship of someone (*tutore auctore Glaucio*) seems to bring it into complete accord with Roman law.

After the name Demetria we may assume a statement of age, as *annis XXXV*, and some distinguishing mark or lack of it. At the end of the lacuna faint traces of letters are seen that might be reconcilable with *facie*, but there is no good reason why the *facie deducta* of the following descriptions should be in inverse order.

c̣um coṃi[te is made somewhat doubtful by the failure of the bottom of the first letter to curve to the right, but a close parallel is seen in the *c* of *cor* at the end of line 3, or in *subcalvo* of line 4. Furthermore, some such connection is demanded by the certain feminine name *Ticol[* , with which compare Τικολλοῦθος, the name of a woman in Maspero, *Papyrus grecs d'époque byzantine*, II, 67158, 8; 11.

2 In the word *ạnn(is)* the facsimile seems to show a dot between *ann* and the sign for *is*, but this is probably only a remnant from the stroke at the top of the second upright of N. It does not appear in the same abbreviation near the end of line 3. A similar abbreviation for *annis* (*ans*) is found in a Christian inscription of 452 A.D. (see Diehl, *Inscriptiones Latinae*, p. 35, No. 5) and in late manuscripts. The description of the attendant is a close parallel to that of the guardian two lines below. Only *lentico* presents difficulty. It seems an adjective formation from the root of *lentigo* and *lenticula* and so means "freckly" or "covered with freckles."

3 *p̣[ac]ṭa ẹ[st]* is a doubtful reading, as only one letter can be considered certain. Yet a verb meaning "has married" or "has betrothed herself" seems required by the connective *que* of *eique* in line 7 and the following verbs *dixit deditque*. The remnants of writing are irreconcilable with *nupsit*, which is expected, and even *nupta est* seems precluded by the curved top of the first letter, which is plainly visible. The verb must govern the dative *C. Valerio Gemello*.

Pannonianus, the father of the guardian Glaucius, must have had some association with the province of Pannonia. I do not find the name recorded elsewhere, though it is a natural derivative from Pannonius. The father was perhaps a soldier or the son of a soldier who was born in or served in that province.

C̣onsọcolẹḍeṭạṛ is the cognomen or second name of Glaucius. It does not seem explainable as Latin, Greek, or Coptic. It joins

the father's name in indicating a northern barbarian origin of the family. Though several of the letters are dotted as somewhat doubtful, in each case enough is visible to warrant the reading given if it produced an explainable result.

4–7 *facie deducta* means "long faced"; cf. *De physiognomonia liber*, 5, *partes inferiores deductae paulatim plenitudine desinente*.

These lines are perfectly preserved so that the chief problem presented is the legal one, which has already been discussed.

6 *Ḷịbyrni:* the spelling seems due to Greek influence; cf. Λιβυρνοί. It is an adjective agreeing with *Dracontis* instead of the common *Liburna* with *navis* understood.

8 This is an example of *dos æstimata* without enumeration of the individual articles of clothing, etc. *Æstimium* is a late word for *æstimatio*. It is recorded only in *Corpus Glossariorum*, Hyginus, *de limitibus constituendis*, and Ps. Boethius, among the Grommatici. Our example seems the earliest recorded use.

In numerato is for *in numerato argento* and therefore means "money" with emphasis on its being coined and so counted; cf. Livy, 36, 21, 11. In Nepos, *Timotheus*, 1, *pecuniam numeratam* is similar and has been regularly interpreted as "cash." However, the addition of *praesens* here, which is used to designate a payment as made at the moment, or in cash, shows that *numeratum* probably puts the emphasis on the fact that the money was counted. In bills of sale the Romans acknowledged the receipt of the payment in good coin accurately counted; cf. P. Lond. 229.

praesens is found in *Digest*, 7. 1, 4, contrasting a cash payment with a payment on time (*quod vel praesens vel ex die dari potest*).

9 In the amount of the dowry only *ḍṛạc̣ṃas* can be considered defensible. So little is preserved of the tops of the letters that one can only say that the preserved remnants are not inconsistent with the attempted restoration.

is Valeri[*u*]*s Ge*[*mellus*]; the use of *is*, perhaps for *hic*, seems unusual and perhaps Vulgar Latin. It may be that in the lacuna *s*(*upra*) *s*(*criptus*) followed the name, in which case *is* would be used as the equivalent of the Greek article. In Latin bills of sale it is usual for the seller to acknowledge that he has received and still has the money in question. Some such addition is expected here.

Signatures of Witnesses

1]ι̣λ̣τ̣ητιος πυλ
Ιηιουνος
2]μ̣ηνλι̣ος βε̄
Τιοφηνος
3]. . .μ̣ι̣ο̣ς̣ γα̣
Ζαδνος
4]α̣λψο̣υιος φα̣λ̄
]στ̣υ̣νος
5]ζ̣ῑ̣ωζιος αρν·
Γην̣ος
6]τ̣ο̣χαυιος κ̣ρι
Μαγγος
7] Αυιλλητος κυρ
Αντωνινο̣ς

There are seven signatures in seven different hands and so presumably the personal signatures of the seven witnesses. As the husband was a soldier in the Alexandrian fleet we may expect the witnesses to be of a similar rank, which does not presuppose great knowledge of language or elegance in writing. Yet the scrawls that pass for signatures are even worse than would be expected. It is true that these witnesses may all be members or former members of the fleet and hence know Latin and Latin script better than Greek. It seems to have been the rule that the signatures must be in Greek. At least only one witness to a marriage contract (the last on Mich. Inv. 508–2217) is known, who signed in Latin. On the other hand there seem Latin cursive tendencies in most of the signatures of this document, and in some cases understandable names can be formed only by assuming a Latin form of a letter rather than a Greek. I do not therefore attempt to explain and justify my reading in each case, but suggest that anyone interested start with the last signatures first. Here we determine that a cognomen stands on a second line in each case. That precludes the thought that military titles are used. Before the cognomen only the tribe, the father's name, or the nomen can stand. The odd combination of consonant and vowel endings of the first lines seems to show abbreviations of the tribal names. Thus πυλ of sig. 1 is for *Pol*(*lia tribu*), βε̄ of sig. 2 is *Ve*(*lina tribu*), γα̣ of sig. 3 is *Ga*(*leria tribu*), φα̣λ̄ of sig. 4 is *Fal*(*erna tribu*), αρν. of sig. 5 is *Arn*(*iensi tribu*), κ̣ρι of sig. 6, is perhaps for *cru*, sometimes used for *Clu*(*stumina tribu*), though it may be careless for κυρι = *Quiri*(*na tribu*), and κυρ of sig. 7 is either *Quir*(*ina tribu*) or *Cor*(*nelia tribu*).

The separation of these tribal designations seems to leave the ending

ιος for the preceding names in most cases, though others look more like ιου. However, as two of the cognomina also seem to end in υ, it has seemed best to read these as ς in all cases and assume influence of the Latin cursive *s*.

In the first signature the ending of the gens name ιλτητιος suggests no parallel, though slight changes would give Ταιήτιος, or some Greek equivalent for *Bitutius* or *Helvetius*. πυλ̣ for *Pollia* is paralleled in Greek; cf. *Pullea* cited with references in Preisigke, *Wörterbuch der griechischen Papyrusurkunden,* III, p. 279. For the cognomen Ιηιουνος the Latin *Ieiunus* = fasting, is the proper parallel.

In the second signature Μηνλ̣ιος is for Manlius, a well known gens name. The form of the b in the tribal abbreviation is regular for the Latin cursive *b* and is found rarely in Greek, but of the second century B.C. or the fourth A.D. Τιο̣φηνος is a pronounceable name, but I find no parallel.

In the third signature little is read of the gens name, but the cognomen Ζαδνος seems clear. Somewhat similar is Ζαβδᾶς in Preisigke, *op. cit.*

The fourth signature presents a problem as one can read]α̣λψι υιος instead of the doubtful gens name]αλψουιος. Were this a Greek such a designation would be usual, but the tribal designation φα̣λ̣ marks the man as a Roman citizen, even though two of the letters are so carelessly written as to be doubtful. As a Roman citizen he must have the three names, and if all the preserved part is taken as the father's name plus ὑιός, there would not be room for both prænomen and nomen before it. For the cognomen]στ̣υνος compare the Greek name Ἀστύνοος in Pape.

In the fifth signature the gens name apparently ended in ζιος. Also a τ seems written above a preceding letter ι as a correction or addition. No parallel occurs to me. For the cognomen Γηνος we may compare Γάνος in Preisigke, *op. cit.*

The sixth signature is the worst written of all. Here also with a little imagination one may read υιος instead of the end of the gens name, but for the reasons given above I assume the gens name. One is tempted to read . . τ̣αυιος and compare Octavius, but χ is more likely than τ and it is certainly preceded by a round letter, as ο. The cognomen Μαγγος seems clear, and stands for the Latin *Magnus.*

In the last signature we certainly read the whole gens name Αυιλλ'ηιος, to which we may compare Ἀυίλλιος in Preisigke, *op. cit.*, and the well known Roman name *Avillius,* which also generally has the double *l.*

The cognomen Αντωνινο̣ς may suggest a date for this document as late as the middle of the second century, but the name, a derivative of

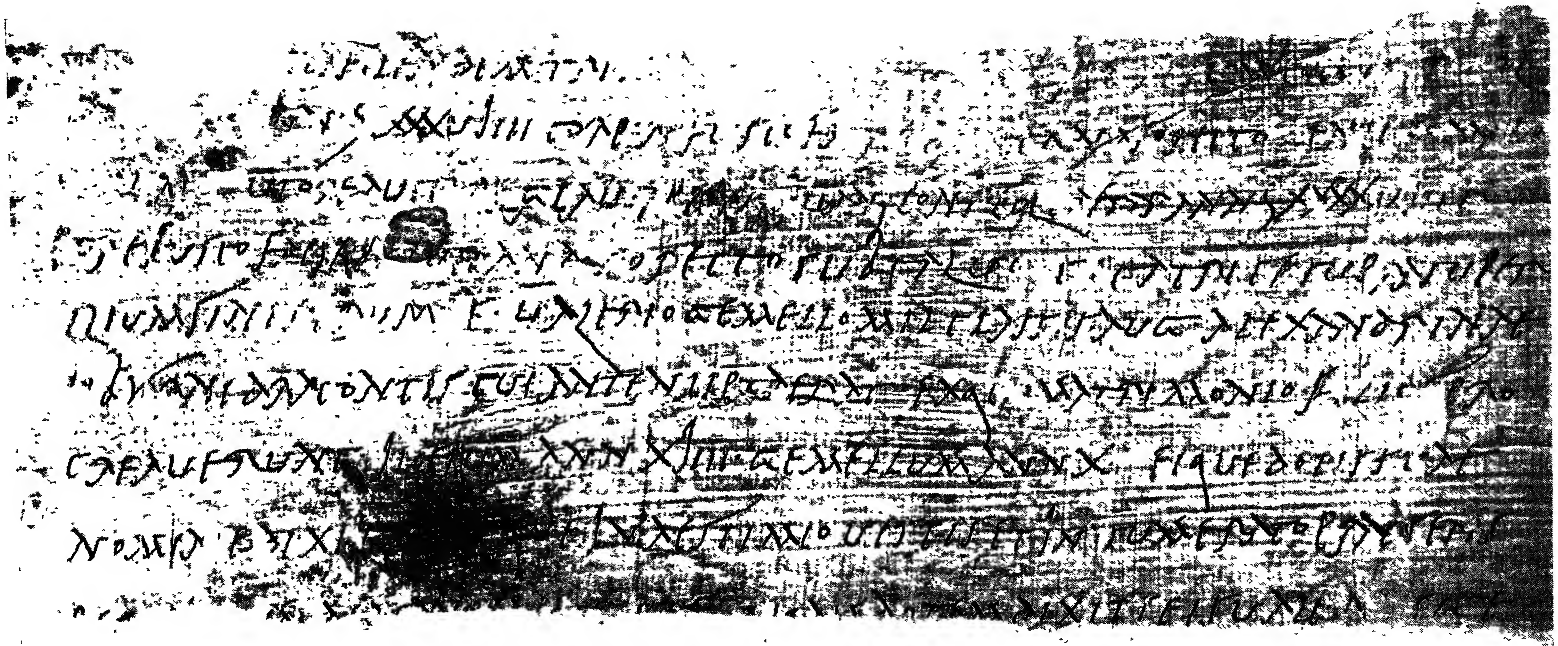
Antonius, is known much earlier; cf. Q. Haterius Antoninus, consul in 53 A.D.

In conclusion it may not seem out of place to use the results of this study on the form of a diploma to interpret a much discussed passage of the New Testament, *Revelation* v. 1, which is translated in our Bible as follows: "And I saw in the right hand of him that sat on the throne a book written within and on the back, close sealed with seven seals." This was a secret book that no one could open or look within; cf. *Rev.* v. 3. Textual difficulty arose early with the substitution of ἔξωθεν for ὄπισθεν, *i.e.* "without" in the place of "on the back." Yet already in the Byzantine period the present interpretation was fixed by placing a comma after ἔξωθεν, "without." As early as 1660 Grotius suggested the removal of the comma and thus read "a book written within and sealed without with seven seals." The reply to this was that the oldest text read ὄπισθεν (on the back) and not ἔξωθεν (without),[1] and that ὄπισθεν suggested ὀπισθόγραφος (written on the back also), a name given to a papyrus roll that was written on both sides. So in spite of the appeal of the Grotius correction, it has been adopted by but two New Testament editors, E. Nestle and T. Zahn.

Now a book at that time would be a papyrus roll and, if written on both sides, a certain portion of the outside writing would be visible, which does not agree with its being a secret book. Also the book seems to have contained the divine judgment, for the world of sin was destroyed, when it was opened. Thus everything suggests to us not a book but a Roman document, sealed and signed on the back by seven witnesses. The marriage diploma is not, however, the best parallel, for that had a copy of the contract also on the outside. Double documents were, however, not common and any document might be sealed, as a will, if it were necessary to keep its contents inviolate or unchanged; see the directions for acknowledging seals, cutting the thread, and opening wills in the Ravenna papyrus (Bruns, *Fontes Juris Romani*, 123). Wills and certificates of sale also seem to have been sealed with seven seals, though not all of the signers are designated as witnesses. The natural interpretation of the Greek text, the chiastic word order, and the parallelism of Roman documents with seals and signatures on the back, all support the Grotius interpretation of *Revelation*, v. 1.

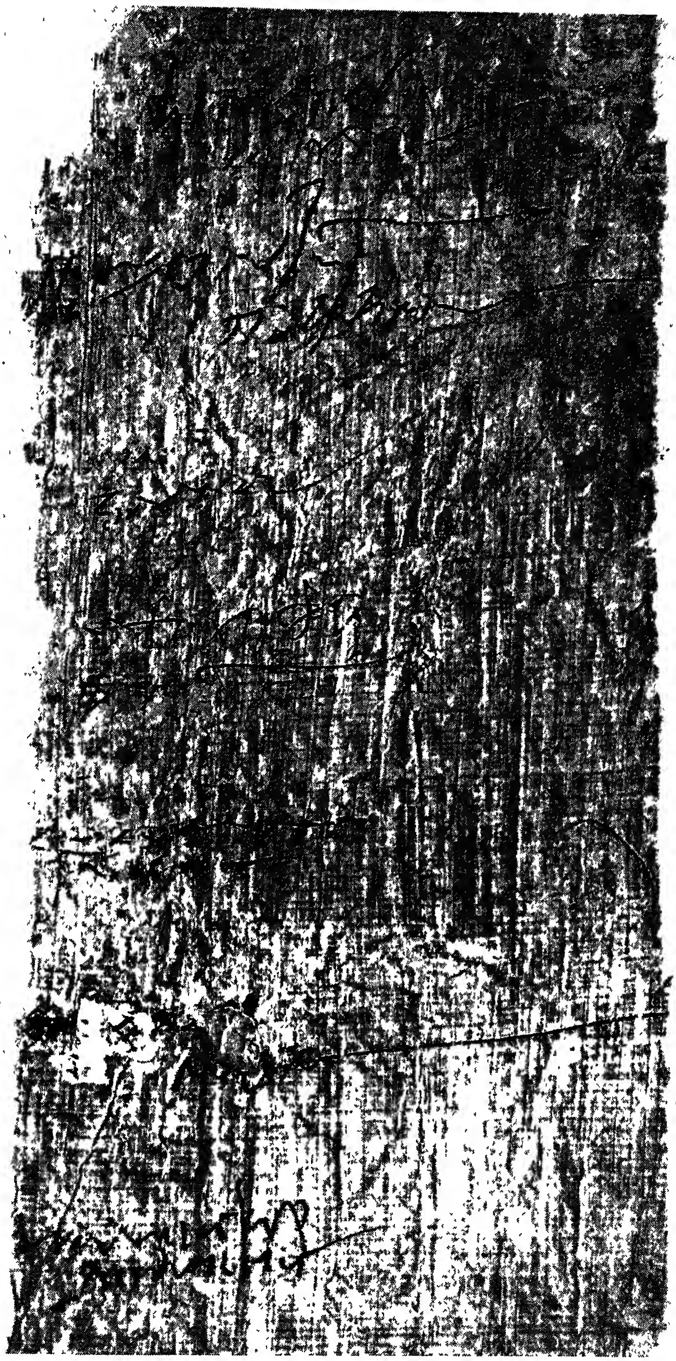
[1] Otto Roller, *Zeitsch. f. d. N. T. Wissensch. u. d. Kunde d. ält. Kirche*, XXXVI (1937), 98–113, defends ἔξωθεν in order to explain the "Book" as a double document of the Greek form.

PLATE I



Marriage Contract.

PLATE II



Signatures of Witnesses.

NARRATIO DE ITINERE NAVALI PEREGRINORUM HIEROSOLYMAM TENDENTIUM ET SILVIAM CAPIENTIUM, A.D. 1189

Edited from the unique manuscript in the Library of the Turin Academy of Sciences

CHARLES WENDELL DAVID,
Professor of European History
Bryn Mawr College

Acknowledgments

It is a pleasure to acknowledge the obligations which have inevitably been incurred in the course of preparing the present work for publication. To the American Council of Learned Societies I am deeply grateful for the assistance which enabled me to visit the south of Portugal in 1934 and examine the site of Silves and see the surrounding country for myself, and also to journey to Italy and work at first hand with the manuscript of the *Narratio de Itinere Navali* in Turin. To Dr. Pericle Maruzzi, Librarian of the Turin Academy of Sciences, I am indebted not merely for permission to use the manuscript under ideal conditions for as long as I wished, but also for his patience in providing me with the photographs which I required and, more important still, for having through a fortunate correspondence informed me of the location of the manuscript before I had seen the description of it which was published by Federico Patetta in 1917. To Senhor Pedro P. Mascarenhas Júdice of Silves I offer my especial thanks for his extreme kindness to me when I visited his interesting city and afterwards and for the generosity with which he placed at my disposal his fund of local historical and archaeological information. To Professors Philip K. Hitti of Princeton University and Solomon L. Skoss of Dropsie College of Hebrew and Cognate Learning I am greatly obliged for timely help in dealing with Arabic matters which I had little competence to handle for myself. To the American Philosophical Society I am obligated, not only for the publication of my manuscript, but for the kind assistance of Dr. Edwin G. Conklin, its Executive Officer, and his efficient staff in seeing the work through press. Finally, I express my very special thanks to my wife who has helped me greatly in the preparation of the manuscript for press, in proof-reading, and in the making of the index.

C. W. David

Bryn Mawr College,
September, 1939.

Contents

Maps and Illustrations

Abbreviations

Chroust .... *Narratio Itineris Navalis ad Terram Sanctam*, in *Quellen zur Geschichte des Kreuzzuges Kaiser Friederichs I* (ed. A. Chroust, Berlin, 1928: *M.G.H.*, *Scriptores Rerum Germanicarum*, new series, vol. V), pp. 179–196.

Gazzera .... *De Itinere Navali, de Eventibus, deque Rebus, a Peregrinis Hierosolymam Petentibus, MCLXXXIX, fortiter Gestis, Narratio* (ed. Costanzo Gazzera, in R. Accademia delle Scienze di Torino, *Memorie*, 2nd series, II, 1840, *Scienze Morali, Storiche et Filologiche*, pp. 177–207).

Idrīsī .... *Description de l'Afrique et de l'Espagne*, par Edrisi (Texte arabe publié pour le première fois d'après les manuscrits de Paris et Oxford, avec une traduction, des notes et un glossaire, par R. Dozy et M. J. de Goeje, [Leyden], 1866).

Kurth .... Friedrich Kurth, *Der Anteil Niederdeutscher Kreuzfahrer an den Kampfen der Portugiesen gegen die Mauren* (in Institut für Oesterreichische Geschichtsforschung, *Mitteilungen*, Ergänzungsband VIII, 1909, 131–252).

M.G.H. .... *Monumenta Germaniæ Historica* (ed. G. H. Pertz and others, Hanover, etc., 1826–).

Silva Lopes, *Memorias* .... Silva Lopes, João Baptista da, *Memorias para a Historia Ecclesiastica do Bispado do Algarve* (Lisbon, 1848).

Silva Lopes, *Relação* .... *Relação da Derrota Naval, Façanhas e Successos dos Cruzados que Parti'rão do Escalda para a Terra Santa no Anno de 1189* (Latin text with Portuguese translation and notes, by João Baptista da Silva Lopes, Lisbon, 1844).

Introduction

Maritime Crusading from Northwestern Europe During the Twelfth Century

The launching of a major crusade to the Holy Land invariably provoked a considerable outpouring of the humbler elements of the population of northwestern Europe who embarked upon the long and perilous, but less costly and more practical, journey to the East by sea. These movements from the North commonly had other objectives than that of a mere pilgrimage to the Holy Sepulchre, and they may have for the modern student an interest in some respects wider than that of a great overland crusade. The commercial or piratical motive is usually recognizable in them; they played a not inconsiderable part in the Portuguese *Reconquista;* and they illustrate the development of northern maritime enterprise along the Atlantic seaboard at a time when our knowledge of it is exceedingly meagre and almost unobtainable in any other connection. Their essential importance requires that they should be made the subject of the fullest possible investigation.

Our knowledge of these movements at the time of the First Crusade is very unsatisfactory, but we are remarkably well informed concerning them in connection with the Second and the Third. We know that many thousands of the northerners sailed for the East in 1147 and again in 1189 and 1190. Indeed, these expeditions were not confined to periods of the great overland crusades, and had we more abundant sources of information, we should perhaps perceive that once the expanding force of northern maritime enterprise had been turned into crusading channels, it hardly ceased to operate so long as the general crusading movement lasted. And from about the time of the Third Crusade the advantages of the sea route to the East began to be so fully recognized that crusaders of high social rank no longer feared or scorned to take it.[1]

In an earlier work I have tried to bring together all that is known concerning the maritime crusading enterprises from the North, their character, and their objectives, down to about 1150.[2] Thereafter for the next generation, or until the time of the Third Crusade, our infor-

[1] Vikings of rank had, of course, been going to the Mediterranean by sea since before the crusading movement began (see Paul Riant, *Les Expéditions et Pèlerinages des Scandinaves en Terre Sainte*, Paris, 1865, *passim*), but it is difficult to find men of social importance from the non-Scandinavian North embarking on crusade by sea before the sailing of King Richard's fleet in 1190. They are much more prominent among the maritime forces of 1217. See Kurth, pp. 215–244.

[2] *De Expugnatione Lyxbonensi* (ed. C. W. David, New York, 1936), pp. 12–26.

mation is painfully inadequate. We are told that in 1150 the new English bishop of Lisbon, Gilbert of Hastings, was back in England preaching a crusade and raising forces for a proposed conquest of Seville.[3] What success he met with we do not know; pretty certainly he never led any crusaders to an attack upon Seville, but it has been supposed that his recruits may have played a part in an unsuccessful attack by Affonso Henriques, the king of Portugal, on Alcácer do Sal.[4] All that we know for certain is that the king took Alcácer, apparently without outside assistance, in June, 1158; but the *Chronica Gothorum* explains that before this he had failed twice to take the stronghold when he had been aided by a great number of ships from the North.[5] It has been suggested that some of these northern ships had been supplied by the adventurous earl Ragnvald of Orkney while on a crusade the history of which is related so uncertainly in the *Orkneyinga Saga.*[6]

From Arabic sources we learn of important naval engagements about 1179 to 1181 between the Almohade fleets of Seville and Ceuta and a Christian fleet of Lisbon.[7] Whether this fleet of Lisbon was partly made up of ships from the north or whether it was wholly the creation of the Portuguese monarchy, we have no means of knowing.

With the launching of the Third Crusade the darkness lifts to a large degree. An Almohade historian, writing of the loss of Silves in 1189, remarks that after Jerusalem had been wrested from the hands of the Christians (1187) it became customary for fleets of crusaders to go every year to the Holy City to fulfil their vows; [8] and from the Christian sources we are able to follow, albeit with some difficulty and confusion, a really great outpouring of northern maritime crusaders in both 1189 and 1190. Of those who seem to have sailed earliest and who reached the coast of Syria early in September, 1189, the conspicuous achievement en route was the conquest of Alvor, on the south coast of Portugal,

[3] John of Hexham, *Historia*, in Simeon of Durham, *Opera Omnia* (ed. Thomas Arnold, London, 1882–85), II, 324.

[4] A. Herculano. *Historia de Portugal* (8th ed. by D. Lopes and P. de Azevedo, Paris and Lisbon, n. d.), III, 65–67; Kurth, pp. 159–161.

[5] *Chronica Gothorum*, in *Portugaliæ Monumenta Historica* ([ed. A. Herculano and others] for the Academia das Sciencias de Lisboa, Lisbon, 1856–1917: incomplete), *Scriptores*, I, 15: "Jam quidem prius obsederat eum per duas vices adjutus multitudine navium que advenerant de partibus Aquilonis, id est de Francia, et finitimis ejus partibus."

[6] *Orkneyinga Saga*, ed. G. Vigfússon with an English translation by G. W. Dasent, in *Icelandic Sagas and Other Documents Relating to the Settlements and Descents of the Northmen on the British Isles* (London, 1887–94), I, 159–179, III, 163–182; Riant, *Expéditions et Pèlerinages des Scandinaves en Terre Sainte*, pp. 254–255; Kurth, 161. Ragnvald and his followers first took a Christian castle in Galicia from its tyrant chief named Godfrey; and then sailed onward and "harried wide in that part of Spain which belonged to the heathen, and got there much goods."

[7] See below, Appendix A, p. 656.

[8] See below, p. 666.

which has been discussed in an appendix to this work.[9] The slaughter of the defenders in which it ended thrilled the Christian world and horrified the Muslim. The expedition of King Richard's fleet of more than a hundred vessels around Portugal and Spain in 1190 is already fairly well known and has been discussed anew in full detail in a Bryn Mawr doctoral dissertation which has recently appeared in print.[10] Elsewhere in this work I have had occasion to note how the timely arrival in Portugal of the earliest of King Richard's forces played an important part in turning the tide against the Muslims at a critical moment in the struggle between King Sancho I of Portugal and the Almohade caliph al-Manṣūr.[11] I have also noted what little is known about the seizure and partial destruction of Silves by a body of maritime crusaders from Germany in 1197.[12]

Of all the northern maritime enterprises which played a part in the *Reconquista* during the twelfth century there are two that stand out above all others, and it is our good fortune that remarkably full accounts of each of them, written by alert and intelligent participants, have survived. The remarkable narrative of the conquest of Lisbon in 1147 I have already edited in an earlier volume under the title *De Expugnatione Lyxbonensi*.[13] Here is presented the account of the conquest of Silves in 1189 from the unique manuscript now in Turin.

The Turin Manuscript and Its Provenance

The historical memoir which is here published under the somewhat cumbersome title *Narratio de Itinere Navali Peregrinorum Hierosolymam Tendentium et Silviam Capientium, A.D. 1189*,[14] has survived in but a single manuscript which is now preserved in the Library of the Turin Academy of Sciences under the press-mark MM. V. 11. In addition to the *Narratio de Itinere Navali* this little volume contains (folios 1r-3v) an important copy of the well-known *Epistola de Morte Friderici Imperatoris* which is probably to be ascribed to Bishop Gotfried of Würzburg.[15] The manuscript was discovered and purchased in 1837 by the Italian scholar Costanzo Gazzera in an outdoor bookstall in Aix-en-

[9] Below, pp. 663–666.

[10] B. N. Siedschlag, *English Participation in the Crusades, 1150–1220*, privately printed, 1939.

[11] Below, p. 658.

[12] Below, p. 660.

[13] New York, 1936.

[14] For the explanation of this title see below, p 610, note 1. For convenience hereafter I shall frequently use the shortened form *Narratio de Itinere Navali*.

[15] The manuscript has been used by Chroust for his recent edition of the letter in *Quellen zur Geschichte des Kreuzzuges Kaiser Friedrichs I* (ed. A. Chroust, Berlin, 1928: *M.G.H.*, *Scriptores Rerum Germanicarum*, new series, V), pp. 173–178, cf. *ibid.*, pp. xcvi–xcix.

Provence.[16] Remaining in his possession until his death in 1859, it passed by bequest with the rest of his library to the Turin Academy of Sciences. There, for want of cataloguing, it was practically lost to view until the publication of an account of the Turin Academy manuscripts in 1917 by Federico Patetta once more brought it to light.[17] Even then it seems almost to have escaped notice until a brief review of Patetta's article by Adolf Hofmeister in *Neues Archiv* (XLIV, 1922, p. 155) gradually brought it to the attention of the scholarly world.[18]

The manuscript (now in a beautiful full-calf blue binding with decoration and lettering in gold, which dates from Gazzera's time) [19] consists of 12 folios of fairly fine vellum, measuring 158 by 103 millimeters, arranged in two gatherings, the first of 8 leaves, the second of 4. Within each gathering the arrangement is in accordance with what E. K. Rand has called Rules I and II,[20] that is, with the hair-side of the parchment facing hair-side, and flesh-side facing flesh-side, and with hair-side used for the outside of both the first and the last leaves. The dirty and worn appearance of these outer pages suggests that the manuscript has suffered from long exposure without the protection of binding or fly-leaves.

The folios are now numbered from 1 to 12 in pencil with arabic numerals in a hand of the nineteenth century (presumably Gazzera's). There are no signatures or early marks of foliation, and nothing indicates that the manuscript ever formed part of a larger volume. All of the leaves are marked with prick-holes near the margin as a guide to ruling, and it is evident that the manuscript was originally ruled with a plummet. Though most of the ruling has now disappeared, particularly from the hair-side of the leaves, there are still faint traces of

[16] Costanzo Gazzera, *Trattato della Dignità ed Altri Inediti Scritti di Torquato Tasso* (Turin, 1838), p. 47: "Nel rovistare, come è mio costume, l'ultimo giorno del mio soggiorno in questa città, i pochi e sudici volumi di un muriciuolaio, mi venne tra mani un quaderno pergameno, e manoscritto, di dodici carte, in-8°, slegato, ed in sì misero stato, che ben indicava aver più d'una fiata dovuto soggiacere agli insulti dell'intemperie, dell'acqua, e del fango; tanto n'erano aspersi e saturi i pochi foglietti che nascondevano in alcune parti il carattere stesso della scrittura. Esaminato, per quanto mi fu permesso, il manoscritto e letto alcune poche linee, la scrittura mi parve d'argomento storico: onde senza più, e per pochi soldi, ne feci l'acquisto." Cf. *idem*, "Narrazione Storica Contemporanea delle Avventure e delle Imprese di Una Flotta di Crociati Partita dalle Foci della Schelda l'Anno MCLXXXIX," in R. Accademia delle Scienze di Torino, *Memorie*, 2nd ser., II (1840), *Scienze Morali, Storiche e Filologiche*, pp. 179–180. On the life and writings of Gazzera see a notice by Casimiro Danna in *Rivista Contemporanea*, XXIV (Turin, 1861), 428–441.

[17] "Di Alcuni Manoscritti Posseduti dalla Reale Accademia delle Scienze di Torino," in R. Accademia delle Scienze di Torino, *Atti*, LIII (1917–18), 543–553.

[18] Hofmeister had sought for it in vain at the time of the publication of his article "Zur Epistola de Morte Friderici" in *Neues Archiv*, XLI (1919), pp. 705–708.

[19] Cf. Patetta, *op. cit.*, p. 550.

[20] *Palæographia Latina*, ed. W. M. Lindsay, V (1927), p. 52.

lines in a light brown color on some of the flesh-side pages, notably folios 7v and 9v.

The manuscript has suffered serious deterioration from exposure and dampness. Several of the leaves are badly discolored and some portions of the text have been rendered illegible. A part of the top margin of folio 12 has broken off and disappeared, carrying a part of a line of text with it. Folios 8v and 9r are spattered with dirt, or possibly with ink of ancient origin. Folios 10v and 11r appear to have suffered from some accident with modern ink.

The space on the pages occupied by writing measures about 135 by 85 millimeters. The number of lines to a page varies from 25 to 33 with great irregularity. Some of the pages are written with considerable spaciousness while others are exceedingly crowded: thus, for example, folio 9v with its 27 lines of easy writing contrasts strikingly with folio 5v or 6r with its 33 lines of close-packed script. In the lower margin of folio 11v there is a rude drawing (possibly by the hand of the original scribe of the manuscript) of the head, neck, and shoulders of a human figure, probably a woman, with parted hair and wearing a headgear which somewhat resembles that of a modern *chef de cuisine* and is ornamented across the front with a row of seven dots.[21] Folio 12r contains only 16 lines, the remainder of the page being blank; folio 12v is blank.

The entire manuscript is written in a single hand which, by the common agreement of competent scholars who have examined it, has been assigned to the beginning of the thirteenth century.[22] Chroust has expressed the view that it was produced in the south of France,[23] but on this point the present editor feels no certainty. Palaeographically it seems to present few peculiarities which call for detailed comment. The writing, while not elegant, is fairly regular. Abbreviation, though extensive and not always regular, generally presents no serious difficulty. Capitalization is fairly common, particularly at the beginning of sentences, and there is a good deal of punctuation by means of single points, which are used indifferently for both full stops and pauses. Some division into paragraphs, or chapters, is indicated by means of rather prominent capitals which are set out a little way into the left-hand margin.

[21] See frontispiece.

[22] Costanzo Gazzera, *Trattato della Dignità . . . di Torquato Tasso*, p. 48; idem, in R. Accademia delle Scienze di Torino, *Memorie*, 2nd ser., II (1840), *Scienze Morali, Storiche e Filologiche*, p. 180; Federico Patetta, *op. cit.*, p. 549; Chroust, p. xcvii.

[23] *Ibid.*, "Die Blätter sind von einer Hand des beginnenden XIII. Jahrhunderts, die ich für eine südfranzösische halte."

From the fact that the manuscript contains two separate works of quite different character and origin, as well as from the numerous unintelligent blunders in its writing, it is perfectly apparent that it is a copy, indeed a very poor copy; yet fortunately most of the mistakes are of so simple and obvious a character as to present few serious editorial difficulties. A notable feature of the manuscript is the evidence it contains of a corrector's hand which is revealed not only in its greater expertness but by the use of a better ink which has defied time and retained its rich dark color while the writing of the original scribe has faded brown. Note has been taken of the corrections in footnotes to the printed text below, so that there is no need for a detailed description of them here. Sometimes they consist of no more than the re-writing of a single letter for the sake of greater clearness; sometimes they take the form of additions above a line or in the margin. They are not always successful or intelligent corrections, as may be seen by referring to p. 614, note 59, below; and many evident errors have been passed over unnoticed.

For the convenience of scholars on this side of the Atlantic who may be interested in the Turin manuscript, the photographic copy of it which I have used will in due course be placed in the Library of Congress, in accordance with the well-known arrangement whereby that institution has been made a national repository for such reproductions of foreign manuscript materials.

The Author and His Work

Whatever the date of the Turin manuscript, it is perfectly certain that the author of the *Narratio de Itinere Navali* was an active participant in the expedition about which he writes and that the composition of his memoir was almost, or quite, contemporary with the events recorded. It has been suggested that his reference to the conquest of Lisbon by northern crusaders "forty-four years earlier" (*ante quadraginta et quatuor annos*) may indicate 1191 as the date of composition;[24] but there can be little doubt that he was reckoning not from the moment of writing but from the time of his arrival in Lisbon, and Chroust [25] has quite rightly pointed out that he has made a mistake.[26] The narrative is not in the form of a diary, and it would be rash to maintain that it was actually written as events occurred; but much of it is in the form of a day to day account and it is difficult to escape the conviction that

[24] Kurth, pp. 164–165.

[25] Page ci.

[26] See below, p. 616 and note 88. The author is in error again in a later passage about the date of the conquest of Lisbon: see below, p. 642 and notes 441 and 442.

in its present form it was composed very early from notes which were made as the expedition progressed. Chroust has pertinently remarked that there are no forward references to later events.[27]

The author must unfortunately remain anonymous, and we know nothing about him except what may be inferred from his work. He was certainly a German, and probably from lower Germany like most of his fellow crusaders who sailed with him from the mouth of the Weser. He writes of himself and his associates as *nos de regno Teutonico*.[28] Arrived in Lisbon, he learns that *naves de nostro imperio et de Flandria* which had preceded him by several weeks had gone on to the destruction of Alvor and a safe passage beyond Gibraltar.[29] He gives the distance from Silves to the sea as a German mile (*miliare Teutonicum*), and in other passages it is clear that this is what he means by *our* miles (*miliaria nostra*).[30] He takes special note of the needless death of two foolish crusaders from Bremen before the serious fighting at Silves began.[31] He observes that the Tagus at Lisbon is as wide as the Elbe near Stade.[32] He compares Silves with Goslar—"not very different in size but having many more houses and fine dwellings, and so girt about with walls and moats that not even a hut could be found outside."[33] Our whole impression is that of a man who was familiar with lower Germany, writing for his fellow countrymen at home.

Chroust,[34] following Kurth,[35] is probably correct in declaring that the author belonged to the clerical order; and yet he reveals surprisingly little evidence of special religious interest or clerical bias. He notes the preaching of the crusade *cum indulgentia apostolicae auctoritatis;*[36] he speaks of the visits of himself and his fellows to the famous relics of Oviedo and to the shrine of St. James at Compostela;[37] he complains of the unfortunate necessity of celebrating Pentecost at sea;[38] he mentions the celebration of early mass at Silves before the launching of formal attacks upon the enemy;[39] he refers occasionally to divine aid

[27] Page cii: "Nirgends finden wir Vorverweise, was gleichsfalls auf gleichzeitige Aufzeichnung deutet."
[28] Below, p. 623.
[29] Below, pp. 616, 617.
[30] Below, pp. 616, 618, 640, 642.
[31] Below, p. 618.
[32] Below, p. 616.
[33] Below, p. 619.
[34] Page ci.
[35] Page 164.
[36] Below, p. 610.
[37] Below, pp. 614, 615.
[38] Below, p. 612.
[39] Below, pp. 621, 622.

vouchsafed in moments of crisis or important decisions;[40] he is interested in the military orders in Spain and Portugal;[41] and he records the elevation of a Flemish priest to be bishop of Silves.[42] While the cumulative effect of these items does probably indicate a clerical interest, it must be acknowledged that there is not one of them for which a pious layman might not well have been responsible. I have noted only one brief phrase of direct quotation from the Bible.[43] There are no references to sermons or miracles or martyrs, or to the building of churches for the burial of the dead;[44] and there is little trace of the kind of superstition which we are likely to associate with ecclesiastical writers of the epoch. Certainly our author, though probably a cleric, was not overly priestly.

His literary style is entirely without rhetorical or other adornment. At the beginning of his narrative he announces an intention to write simply: *simpliciter explicare decrevi*,[45] he says, and he has certainly done so. As he indulges in almost no Biblical quotation, so he displays no familiarity with classical authors or with the church fathers and makes no parade of theological learning.[46] He is capable of writing very simple Latin narration with something approaching correctness,[47] but his education must have been rudimentary.

His moral principles are those of the feudal age. He has great admiration for deeds of bravery, particularly individual bravery, and it is an admiration which is not confined to members of his own group:[48] his finest tribute is paid to a Gallegan knight who singled-handed pulled a stone from the corner of one of the towers of Silves right under enemy fire.[49] He has great respect for contracts and is unsparing of criticism when they are violated even by his own associates.[50] He displays a mild anti-Flemish bias,[51] but his principal criticisms are reserved for the Portuguese. He gives King Sancho and his forces mild

[40] Below, pp. 619, 624, 630.

[41] Below, pp. 630, 631.

[42] Below, p. 633.

[43] Below, p. 622, and note 174; others may well have escaped me.

[44] In these matters the present work contrasts strikingly with the closely analogous *De Expugnatione Lyxbonensi*, ed. David, *passim*.

[45] Below, p. 610.

[46] Cf. Chroust, p. ci.

[47] Many of the mistakes in the Turin manuscript are doubtless due to the copyist rather than to the author.

[48] Below, pp. 621, 623, 624, 625, 626, 627.

[49] Below, p. 624.

[50] Below, p. 628: "populus noster satis turpiter quosdam exspoliavit contra pactum et verberavit"; *ibid.*, p. 629: "quidam etiam contra pactum torquebantur pro pecunia monstranda"; cf. *ibid.*, pp. 631, 632.

[51] Below, pp. 623, 632.

credit for their efforts in the earlier stages of the siege of Silves,[52] but after they had repeatedly manifested a disposition to abandon the undertaking and retire, and after a series of disputes involving concessions from the original compact and not unnatural misunderstanding, he ends by roundly condemning the king for shabbiness and for failure to fulfil a vow and by declaring that the Portuguese had neither labored nor fought throughout the whole siege but had only taunted the crusaders for foolishly engaging in a vain effort to take an unexpugnable fortress.[53] Doubtless he was to some extent unfair to his allies, but in the absence of any statement of the Portuguese side of the controversy, it is impossible to say to what extent his strictures may or may not have been justified. He has great respect for the fighting qualities of the Muslims,[54] and there is perhaps a trace of pity in his description of their condition as they went out of the city after their terrible defeat.[55] But he is, of course, a realist about the fortunes of war: it was right for the enemy to submit to the terms of the capitulation, because they were dying of thirst and the mines threatened them and their defences were in a state of ruin.[56]

Chroust has remarked that he betrays himself as a landsman by his somewhat naïve interest in the experiences of sea travel which would be taken for granted by a more experienced voyager.[57] He almost never fails to record the character of the weather when he is at sea; he comments on a manifestation of St. Elmo's fire one terrible night in the Bay of Biscay,[58] and on the tacking of the ships to get through the Strait of Gibraltar against adverse winds,[59] and on the passage of a school of dolphins.[60] But he has the intelligent traveler's alert interest in almost all novelties. He admires the fine houses of Silves and Cádiz,[61] and is much impressed with the walls of mud and plaster and the tile roofs which effectively resisted the crusaders' efforts to burn them.[62] He not only describes the fortifications of Silves in detail,[63] and those of Cádiz with less fulness,[64] but he explains that cities in general in Spain, whether

[52] Below, pp. 619, 623, 624, 625, 626, 627–628.
[53] Below, pp. 626, 627, 628, 629–630, 631–632.
[54] Below, pp. 622, 623, 624, 625, 626, 627–628.
[55] Below, p. 629.
[56] Below, p. 628.
[57] *Op. cit.*, p. ci.
[58] Below, p. 613.
[59] Below, p. 639.
[60] Below, p. 613.
[61] Below, pp. 619, 639.
[62] Below, p. 622.
[63] Below, pp. 619–621.
[64] Below, p. 639.

Muslim or Christian, are fortified according to a common plan, and he makes the correct observation that such strange words as *rovalle*, *almadina*, *corrasce*, *alcaz*, and *alverrana* by which the various parts of a fortress are designated, are common, not proper, nouns.[65] He is also interested in merchants and commerce [66] and remarks upon the opulence of La Rochelle, Lisbon, and Seville,[67] and on the thrice-yearly fair which was held at Cádiz by the Saracen merchants of Africa and Spain,[68] and on Ceuta, the most opulent city of Barbary, *ad quam confluunt omnes Christiani mercatores in Africam commercia transferentes, et maxime Ianuenses et Pisani.*[69]

The author is fond of dates and statistics.[70] He gives a great many dates and, so far as it is possible to judge, they are remarkably accurate. He gives very exact, and apparently correct, information about numbers of ships.[71] He is also very definite when he comes to deal with larger figures: 5,600 pagans had been slain at Alvor, *sicut veraciter audivimus;* [72] there had been four hundred and fifty Christian captives in Silves at the beginning of the siege, but only two hundred remained alive at the end; [73] the total population of Silves, *promiscui sexus*, was 15,800; [74] the total number of crusaders of every rank and age when they first arrived at the siege was 3,500, *vel paulo pauciores.*[75] Our confidence in his veracity is considerably strengthened by the fact that this last figure, which we are able to check with some degree of assurance,[76] appears to be substantially correct.

Apart from the detailed record of the day to day progress of the expedition, and especially of the siege of Silves, the greatest value of the *Narratio de Itinere Navali* unquestionably arises from the author's extraordinary interest in geography. His work fairly bristles with place-names. He records the principal places past which he sailed all the way from Lowestoft to Marseilles; he also lists African place-names from Azemmur below Casablanca all the way to Ceuta; [77] and his

[65] Below, p. 621.
[66] Below, pp. 639, 640, 642.
[67] Below, pp. 612, 616, 635.
[68] Below, p. 639.
[69] Below, p. 640.
[70] Cf. Chroust, p. ci.
[71] Below, pp. 610, 616, 617.
[72] Below, pp. 616–617.
[73] Below, p. 629.
[74] *Ibid.*
[75] Below, p. 630. In the same place he indicates that the Portuguese forces were very strong in cavalry, foot-soldiers, and galleymen, but he gives no figures, presumably because he lacked definite information.
[76] See below, p. 630, note 297.
[77] Below, pp. 637–638, 640.

interest extends inland to such cities as Seville and Cordova,[78] and even to far-away Marrakesh, the capital of the Almohade empire.[79] He notes many scattered facts of geographical interest such as the character of the coastal regions past which he sailed,[80] the dimensions of the Strait of Gibraltar,[81] and the usual place of crossing between Africa and Spain.[82] Occasionally, as Chroust has observed,[83] he interrupts his narrative for a little geographical excursus: as, for example, when he pauses to comment on the nine ecclesiastical sees of Brittany, "three of which use the Breton tongue, a speech which is common to no other people, while the rest share the Gallic idiom," [84] or to describe "the five kingdoms of Spain,"[85] or to make notes on the situations of Lisbon, Silves, and Cádiz,[86] or on the racial geography of southern Spain and northwest Africa,[87] etc. His statements are not always accurate: he makes mistakes when dealing with facts of history, or of distant geography, occasionally even when dealing with matters close at hand.[88] But his errors may fairly be said to be of minor importance, and on the whole his information is remarkably sound. There is no evidence of any dependence on the outworn works of such ancient writers as Pliny and Solinus.[89] One gains the impression that he acquired his information from sailing directions actually in use by navigators or from the sound local geographical knowledge of the lands through which he travelled, much of which was doubtless ultimately derived from the Arabic geographers.[90]

Of the organization of the expedition for purposes of direction, the maintenance of discipline, and the distribution of spoils, the author tells us nothing directly, but from his chance remarks it is possible to conclude that it must have been very similar to that of the Lisbon crusade

[78] Below, pp. 635–636.
[79] Below, p. 638.
[80] Below, pp. 615, 635–638, 641.
[81] Below, p. 640.
[82] *Ibid.*
[83] Page ci.
[84] Below, p. 613.
[85] Below, p. 614.
[86] Below, pp. 616, 619–621, 639.
[87] Below, p. 621.
[88] For example, he represents northwest Africa all the way from the Strait to Marrakesh as a vast plain (pp. 637–638); he is in error twice (pp. 616, 642) about the date of the conquest of Lisbon; he regards Silves as the capital of a kingdom (p. 619), and speaks of the Almohade governor as its king (p. 622), though he later refers to him as *dominus* (p. 628); and he almost certainly exaggerates in describing Silves as more strongly fortified and ten times richer in buildings than Lisbon (p. 629).
[89] Here he contrasts favorably with the author of the *De Expugnatione Lyxbonensi*, ed. David, pp. 38, 67, 86, 87, 90–94.
[90] He commonly gives distances from one city to another in terms of a day's journey, as the Arabic geographers usually did.

of 1147, concerning which we have better information.[91] Clearly the forces were not all of uniform rank,[92] but also quite certainly they were drawn for the most part, if not entirely, from the middle and lower classes. Not a single leader is named, and there is no indication that any one of noble rank had any part in the enterprise. No one is mentioned above the rank of a commoner except a knight of Galicia who had joined the northerners as the pilot or commander of one of their vessels.[93] That these men were bound together in some kind of a formal association, that they arrived at important decisions through common deliberation, and that they endeavored to enforce decisions and maintain discipline by means of their own magistrates, there can be no doubt. The author repeatedly refers to the participants in the expedition as associates (*socii*),[94] and he speaks of a galley from Tuy in Galicia as having joined their association (*quae nobis contubernio iuncta est*).[95] The chosen leaders he occasionally calls magistrates (*magistri* or *magistratus*).[96] More than once he refers to decisions which are reached through deliberation in common,[97] and he also mentions a council in which such decisions are made.[98] Decisions were by no means always unanimous; indeed, sharp disagreements must have been frequent.[99] The orderly and equitable distribution of spoils[100] was, of course, a common and necessary feature of all such enterprises.

The Text

In preparing the present text for publication I have endeavored as a rule to depart as little from the manuscript as possible. Fortunately in the matter of capitalization and punctuation[101] it has been possible to approach an acceptable modern standard without very great variation from the practice of the manuscript, though, of course, strict adherence to it has been out of the question. For the present division of the work into paragraphs the editor is alone responsible, since the few divisions

[91] See *De Expugnatione Lyxbonensi*, ed. David, pp. 56–59.

[92] Below, p. 630: "Noster exercitus tantum habebat . . . tria milia et quingentos cuiuslibet ordinis vel etatis viros."

[93] Below, p. 624.

[94] Below, pp. 612, 613.

[95] Below, p. 617.

[96] Below, pp. 621, 632.

[97] Below, p. 626: "nostri decreverunt communiter diutius hostes Christi impugnare, et hoc regi intimaverunt"; *ibid.*, p. 633: "communem assensum extorquere non potuit;" cf. also pp. 618–619, 627, 628, 631–633.

[98] Below, p. 621.

[99] Below, pp. 623–624, 630.

[100] Below, p. 633.

[101] See above, p. 597.

which are indicated in the original text [102] are quite inadequate to serve the requirements of modern convenience. In the spelling of proper names I have been careful to follow the manuscript exactly, except in one case where the conjunction of a proper noun with a preposition has resulted in an obvious corruption.[103] Apart from proper names the manuscript spellings have been retained except in the case of an obvious slip or where the departure from normal was so extreme as to cause real inconvenience. In such cases the spelling has been regularized in the printed text and the manuscript reading has been given in a footnote. Thus, to illustrate, the word *opidum* is spelled with a single *p* in the printed text; and the diphthong *ae* is never used, since it is the regular practice of the manuscript to substitute the simple *e* for it. On the other hand, I have not hesitated to change *iusta* to *iuxta* or *fondibus* to *frondibus*, or to straighten out such a scribal monstrosity as *celebratissimarum* by printing *celebratis missarum;* I have occasionally corrected grammatical errors by substituting *princeps* for *principes*, *anchoras* for *anchoris*, and the like; and I have avoided such awkward spellings as *hostium* (mouth) and *hamenissimas* by printing *ostium* and *amoenissimas*. There is room for disagreement with such decisions, which are necessarily somewhat arbitrary; but since the true reading of the manuscript, if not printed in the text, is always to be found in a footnote, no serious difficulty can arise.

The manuscript occasionally presents peculiar difficulties which have forced a resort to more serious conjectural emendation. Most frequently these arise from the careless omission of a single word, which can as a rule be supplied with a fair degree of assurance. Where this has been possible, I have made the necessary addition and indicated its presence in the text by enclosure in square brackets. But where there was great or complete uncertainty as to what was wanting, I have simply inserted square brackets in the text to indicate an omission without attempting to fill it.[104] Often difficulty has arisen, not from carelessness on the part of the scribe, but from the present decayed or illegible condition of the manuscript. Here, in cases where I have been reduced entirely to conjecture, or to complete dependence on the readings of Gazzera, I have again used square brackets to enclose the text supplied; but in a few cases where the readings, though difficult and doubtful, were still not wholly conjectural, I have indicated the difficulty merely by the use of italic type.[105] To anyone who examines my notes it will be

[102] See above, p. 597.
[103] Below, p. 610, note 4: *assaahadino* for *a Saahadino* (Saladin).
[104] For example, after *persecutionis* on p. 614, or after *diete*, on p. 635.
[105] Italic type has also been used occasionally for the extension of an uncertain abbreviation

quickly apparent to how large an extent I have profited by the earlier work of Gazzera and Chroust in the matter of conjectural emendation; but I have made it a rule not to resort to it unless certainly necessary; and I have refrained from adopting some of the more extreme conjectural emendations that have been proposed by Chroust and others.[106]

Previous Editions

The *Narratio de Itinere Navali* was first published soon after his discovery of the manuscript by Costanzo Gazzera under the title *De Itinere Navali, de Eventibus, deque Rebus, a Peregrinis Hierosolymam Petentibus, MCLXXXIX, Fortiter Gestis Narratio*, in R. Accademia delle Scienze di Torino, *Memorie*, 2nd series, II (1840), *Scienze Morali, Storiche e Filologiche*, pp. 191–207.[107] The editor introduced it with an essay in which he set forth briefly such facts as he could infer concerning the author and in which he gave a fair summary of the contents of the narrative, though he made an unfortunate mistake in representing the crusading fleet as sailing from the mouth of the Scheldt instead of from the mouth of the Weser. His text was at many points incorrect or incomplete, and in his notes he attempted to solve few of the editorial problems except the identification of place-names, and here he made a good many errors. In short, Gazzera's edition, while its appearance was an event of much importance in that it brought a new source to light, was the work of a scholar who was not a specialist in the period or the problems with which he had to deal, and it left much to be desired; and the cumbersome and expensive series in which it appeared prevented it from being widely circulated and becoming well known.

Gazzera's edition was soon reproduced separately with a parallel Portuguese translation for the Lisbon Academy of Sciences by João Baptista da Silva Lopes under the title *Relação da Derrota Naval, Façanhas e Successos dos Cruzados que Parti'rão do Escalda para a Terra Santa no Anno de 1189*, Lisbon, 1844. Not only was Gazzera's text translated into Portuguese, but also most of his notes and a good part of his introduction; and to the whole Silva Lopes added a considerable body of notes of his own, which were helpful upon some points, particularly the identification of certain place-names where Gazzera had gone astray; but Silva Lopes worked without access to the manuscript so that most

[106] See below, p. 613 and note 37, p. 620 and note 156.

[107] Gazzera did not publish the *Epistola de Morte Friderici Imperatoris* which was already in print from another manuscript. Gazzera's work was reviewed by Le Baron de Reiffenberg in Académie Royale des Sciences et Belles-Lettres de Bruxelles, *Bulletins*, 1st series, VII, pt. 2 (1840), pp. 22–30.

of Gazzera's errors were inevitably repeated and this edition also was far from satisfactory.

After Federico Patetta had once more brought the manuscript to light, it was loaned in Germany by the courtesy of the Turin Academy of Sciences and aroused interest in circles connected with the *Monumenta Germaniae Historica;* and in due course a new edition under the title *Narratio Itineris Navalis ad Terram Sanctam* was included by A. Chroust in his *Quellen zur Geschichte des Kreuzzuges Kaiser Friedrich I* (Berlin, 1928: being *M.G.H.*, *Scriptores Rerum Germanicarum*, new series, V), pp. 179–196.[108] This recent edition by one of the leading palaeographers and editors of Germany should have been satisfactory in all respects, but unfortunately it was not. Chroust had himself examined the Turin manuscript in Würzburg, presumably in connection with his edition of the *Epistola de Morte Friderici Imperatoris;* but for the preparation of his text of the *Narratio de Itinere Navali* itself he seems to have been mainly dependent on a collation of the manuscript which was made by Dr. Gerhard Laehr[109] and he pretty certainly cannot have worked seriously from the manuscript itself or from a photographic copy of it. The result of this unfortunate circumstance is that numerous errors, some of them serious and easily avoidable, have found their way into Chroust's text.[110] Furthermore, since the narrative breaks off abruptly at Marseilles in the autumn of 1189, it really has but the slightest connection with the crusade of the emperor Frederick Barbarossa and hardly belongs in Chroust's collection at all. Being concerned mainly with events in Portugal and Spain and along the Atlantic seaboard, the editorial problems which it raises involved a part of the world with which Chroust was evidently not familiar and in which, it seems fair to say, he must have been but little interested. Hence he has failed to give them the attention which their essential importance warrants, and too often he has been content to rely on Gazzera or on the more recent work of Kurth which, though valuable, was done entirely from the edi-

[108] See also Chroust's *Einleitung*, pp. xcvi–cii.

[109] Chroust, p. xcvii: "Die Hs. ist im Frühjahr 1927 von Dr. Gerhard Laehr in Berlin für eine Neuausgabe sorgfältig verglichen worden"; *ibid.*, p. c, note 4: "Ich konnte für den folgenden Text neben der Hs. die schon erwähnte Vergleichung von Dr. Gerhard Laehr benutzen"; *ibid.*, p. civ: "Auch Herrn Dr. Gerhard Laehr in Berlin sei für die Überlassung seiner Kollation der Turiner Handschrift der Narratio itineris navalis noch einmal gedankt."

[110] As an indication that this is not a frivolous criticism, the following examples of the more serious of Chroust's errors may be cited: below, p. 616, line 9, for the text *processerant nos ante IIII. ebdomadas, vel V., naves de nostro imperio,* he read *processerant nos ante quatuor ebdomadas quinquaginta quinque naves de nostro imperio;* p. 626, lines 8–9, for *diutino et instanti labore nisus est,* he read *diutino et instanti labore visus est;* p. 629, line 3, for *ex alia parte intraverunt nostri, et quidam primo per eandem portam,* he read *ex alia parte intraverunt nostri et quidam Portugalenses per eandem portam;* p. 635, line 13, he emended *Silvia* to read *Sivilia,* thereby confusing a good clear statement.

tions of Gazzera and Silva Lopes without the knowledge that the Turin manuscript so much as existed.[111] It is, therefore, necessary to say that this latest edition, in the justly admired *Monumenta Germaniae Historica,* is also unsatisfactory.

Bibliographical Note [112]

Apart from the Turin text the sources which throw any light on the Silves expedition are meagre indeed. Much the most valuable from the Christian side is the short account which the English chronicler, Ralph de Diceto, has given in his *Ymagines Historiarum* (*Opera Historica,* ed. William Stubbs, London, 1876, II, 65–66). Even briefer, but good as far as it goes, is the notice in the *Gesta Regis Henrici Secundi* (ed. William Stubbs, London, 1867), II, 89–90, which is repeated with small additions and one omission in the *Chronica* of Roger of Howden (ed. William Stubbs, London, 1868–71), III, 18. As I have pointed out below,[113] the reference to the conquest of Silves in the *Itinerarium Peregrinorum et Gesta Regis Ricardi* (ed. William Stubbs, London, 1864), p. 65, probably arises from an erroneous substitution of Silves for Alvor. The most important northern continental source which mentions the conquest of Silves is a brief narrative which I have quoted below from Robert of Auxerre,[114] but which is to be found in several other chronicles. Though somewhat the same confusion exists here as in the *Itinerarium* just mentioned, this has some real value for the Silves expedition. The spurious letter of Pope Clement III to the Byzantine emperor Isaac Angelus, which has often been accepted as a genuine source for the conquest of Silves (even by Kurth and Chroust), was in all probability ultimately derived from this narrative.[115] The only valuable Portuguese source to make any mention of the Silves expedition is the *Chronicon Conimbricense* (*Portugaliae Monumenta Historica, Scriptores,* I, 3) which gives the date of the conquest correctly.

Of Arabic sources which throw any light on the conquest of Silves, much the most valuable are two early Almohade histories, viz.: *El*

[111] See below, note 112; p. 609.

[112] For older or brief indications of the sources, other than the *Narratio de Itinere Navali,* bearing on the Silves expedition of 1189 see Gazzera's introduction in R. Accademia delle Scienze di Torino, *Memorie,* 2nd series, II (1840), *Scienze Morali, Storiche, e Filologiche,* p. 181; A. Herculano, *Historia de Portugal,* 8th ed., III, 342–348; Reinhold Röhricht, *Beitrage zur Geschichte der Kreuzzuge* (Berlin, 1874–78), II, 200–201; Chroust, p. cii. The first full discussion was by Friedrich Kurth in his essay entitled *Der Anteil Niederdeutscher Kreuzfahrer an den Kämpfem der Portugiesen gegen die Mauren,* in Institut für Oesterreichische Geschichtsforschung, *Mitteilungen,* Ergänzungsband, VIII (1909), pp. 164–170.

[113] Appendix B, p. 664, note 4.

[114] Appendix B, pp. 664–665.

[115] See below, appendix B, p. 665.

Anónimo de Madrid y Copenhague (ed. A. Huici, Valencia, 1917), pp. 60–61; and 'Abd al-Wāḥid al-Marrākushi, *Histoire des Almohades* (French translation by E. Fagnan, Algiers, 1893), pp. 243–244. Also important are Ibn al-Athīr, *Annales du Magreb et de l'Espagne* (French translation by E. Fagnan, Algiers, 1898), pp. 608–609; and Ibn Khaldūn, *Histoire des Berbères et des Dynasties Musulmanes de l'Afrique Septentrionale* (French translation by Le Baron de Slane, new ed. by Paul Casanova, Paris, 1925–27), II, 212. Less important, but still perhaps not to be wholly ignored, are the later Arabic works of Ibn Abī Zar', *al-Ḳirṭās* (*Annales Regum Mauritaniae*, edition and Latin translation by C. J. Tornberg, Upsala, 1843–46, II, 190–191); and al-Makkarī, *Nafḥ al-Ṭīb* (Pascual de Gayangos, *History of the Mohammedan Dynasties in Spain*, London, 1840–43, II, 320).

Though the conquest of Silves in 1189 naturally occupied a considerable place in the older secondary histories of Portugal,[116] Herculano was the first historian of note who was able to deal with it in a large way on the basis of a full use of the *Narratio de Itinere Navali* which Gazzera had published shortly before his work began to appear.[117] The new source was also used extensively by Reinhold Röhricht in his *Beiträge zur Geschichte der Kreuzzuge* (Berlin, 1874–78), II, 170–177. But it was not fully exploited until the publication in 1909 of Kurth's essay on the part played by Lower German crusaders in the struggles of the Portuguese against the Moors.[118] Not only did Kurth deal more fully with the subject-matter of the *Narratio de Itinere Navali* than had ever been done before; but, realizing the inadequacies of the then existing editions, he devoted himself to the solution of a number of problems that were properly editorial and so placed both Chroust and the present editor in his debt.

For a complete list of all works, both primary and secondary, which have been cited in the present volume the reader is referred to the index which will include references to all first, or full, citations.

[116] See, for example, Antonio Brandão, in Bernard de Brito, *Monarchia Lusytana*, ([Alcobaça] and Lisbon, [1597]–1727), IV, folios 10–15; Heinrich Schäfer, *Geschichte von Portugal* (Hamburg, 1836–54), I, 104–108.

[117] A. Herculano, *Historia de Portugal*, 1st ed., Lisbon, 1846–53.

[118] Kurth, pp. 164–208.

NARRATIO DE ITINERE NAVALI PEREGRINORUM HIEROSOLYMAM TENDENTIUM ET SILVIAM CAPIENTIUM, A.D. 1189[1]

Antiquorum provide consuetudini[2] morem gerens qui gesta sua scripture laqueis innodare satagerunt ut posteritatis[3] noticiam non evaderent, itineris navalis multiformes eventus qui peregrinis Ierosolimam tendentibus acciderunt simpliciter explicare decrevi.

Anno siquidem dominice incarnationis $\overset{o}{\text{M}}.\overset{o}{\text{C}}.\text{LXX}\overset{o}{\text{X}}.\text{VI}\overset{o}{\text{I}}.$, a Saahadino[4] rege Egypti destructa terra promissionis, captis urbibus, captivatis vel necatis incolis, predicationis tuba cum indulgen*t*ia[5] apostolice auctoritatis[6] late per Christianorum terminos evagata, ad restaurationem miserabilis cladis innumerabilem movit [4r.] populum. Inter quos quibusdam placuit per longissimos tractus maris peregrinationis incolarum[7] pro abolitione criminum erumpnosam semitam[8] protelare.

A Brema[9] autem undecim navibus bellatoribus, armis et cibariis

[1] The work is without title in the manuscript. The title assigned to it by Gazzera (*De Itinere Navali, de Eventibus, deque Rebus, a Peregrinis Hierosolyman petentibus, MCLXXXIX, fortiter gestis, Narratio*) seems excessively cumbersome and is less informing than it should be. Silva Lopes, in his Latin-Portuguese edition, practically translated Gazzera's title, but with an unfortunate addition (for which Gazzera himself was ultimately responsible) containing the misinformation that the fleet set sail from the River Scheldt. The title of Chroust (*Narratio Itineris Navalis ad Terram Sanctam*), while commendably brief, is unfortunately somewhat misleading and contains no reference to the principal event of the expedition, namely the conquest of Silves. It is hoped that the title here adopted is sufficiently like the others to avoid too great bibliographical confusion, while giving some indication of the main subject of the narrative.

[2] MS. *consuetudi.*

[3] MS. *posteritas* was written first and later corrected by the addition of *ti* above the line in different ink.

[4] MS. *assaahadino,* combining noun and preposition. Saladin (1138–1193), sultan of Egypt, conqueror of Jerusalem and the Holy Land.

[5] MS. possibly *indulgencia.* The plenary indulgence of crusaders, first introduced by Pope Urban II at the Council of Clermont in 1095 and soon to become "an established rule in all the holy wars in which the church engaged." H. C. Lea, *A History of the Inquisition of the Middle Ages* (New York, 1888), I, 42; cf. *idem, A History of Auricular Confession and Indulgences* (Philadelphia, 1896), III, 9–10, 152–154.

[6] MS. *actoritas.*

[7] Chroust, feeling that the words *peregrinationis incolarum* do not make sense, has emended the text to read *peregrinando suorum.*

[8] MS. *erumpnosa semita.*

[9] Bremen, on the Weser, some forty-six miles from the North Sea. The reading of the manuscript is certain, although Gazzera unfortunately failed at this point and, in place of *A Brema,* printed *Ab . . .* and stated in a note that the word was obliterated. Kurth (pp. 166–167), without having seen the manuscript, shrewdly conjectured that *A Bremensibus* should be the reading. More recently Federico Patetta ("Di Alcuni Manoscritti posseduti dalla Reale Accademia delle Scienze di Torino," in R. Accademia delle Scienze di Torino, *Atti,* LIII, 1917–18, p. 552) has, with the manuscript before him, perceived that *A Brema* ought to be the correct reading, though he was not wholly successful in seeing this in the manuscript. Chroust has read correctly.

sufficienter instructis, anno dominice incarnationis $\overset{o}{\text{M}}$.$\overset{o}{\text{C}}$.LXX$\overset{o}{\text{X}}$.VIII$\overset{o}{\text{I}}$., [X ?][10] Kal. Maii, de Bleclrente[11] hora nona iter movimus. Sed sequenti die unam navim in arena herentem post nos reliquimus. Nos autem velificavimus, et VII$\overset{o}{\text{I}}$. Kal. Maii[12] in Angliam venimus ad locum qui dicitur Lothevigestohet.[13] Postera die cum tempestate Tense[14] transsivimus, et portum Sanduuinc[15] minus caute intrantes, tres naves ex collisione super arenas perdidimus, salvis rebus et hominibus, quarum due prorsus perierunt, tercia reparata est.

Ibidem moram facientes .XX.III. dierum[16] infra quas navim quam reliqueramus salvam recepimus. Ibi et alie ad nos venerunt, sed pro diversis necessitatibus quedam precesserunt, quedam tardius subsequute sunt.[17] Nostra autem rate perdita, in Lundonia navim comparavimus et, redintegratis utensilibus,[18] $\overset{o}{\text{X}}$IIII. Kal. Iunii[19] a portu

[10] There is an obvious error of omission in the manuscript at this point, since the fleet reached Lowestoft in England on the 8th Kalends of May. Kurth (p. 176, note 4) conjectured that the scribe, having written *VIIII* as a part of the year-date, inadvertently failed to repeat it as a day-date before *Kal.* But it seems hardly likely that a fleet sailing from the mouth of the Weser in mid-afternoon on April 23, and being not improbably delayed by the misadventure that overtook one of the vessels, would have reached England on the 24th. A more probable conjecture, though still only a guess, would perhaps be that *X* should be supplied, thus making April 22 the date of departure.

[11] Probably Blexen on the left bank of the Weser, opposite the modern Bremerhaven. Gazzera, who read *Bleclerente,* identified the name with the Island of Walcheren at the mouth of the Scheldt. But the correctness of this identification, though accepted by Silva Lopes (*Relação*) was questioned in a review of Gazzera's work by Reiffenberg (Académie Royale des Sciences et Belles-Lettres de Bruxelles, *Bulletins,* VII, 1st series, pt. 2, 1840, p. 25) who proposed Flushing instead; and more recently Patetta (*loc. cit.* in note 9 above) has challenged it again, pointing out that if the reading *A Brema* at the beginning of the sentence be correct, the place in question must be far removed from the mouth of the Scheldt. Meanwhile Kurth (pp. 165–167), working independently and without having seen the manuscript, has advanced fairly convincing reasons for identifying *Bleclrente* with Blexen, although Chroust seems to imply some doubt as to the correctness of this identification.

[12] April 24.

[13] Lowestoft, Suffolk, at the mouth of the Waveney, at the easternmost point of England.

[14] The River Thames.

[15] Sandwich, Kent, one of the Cinque Ports.

[16] Presumably April 26 to May 18.

[17] Cf. *Gesta Regis Henrici Secundi* (ed. William Stubbs, London, 1867), II, 89–90: 'Eodem anno [1189], mense Septembris, homines Lundonienses et caeteri multi qui de diversis regnis per naves iter Jerosolimitanum arripuerant, obsiderunt in Hispania civitatem quandam Saracenorum quae Silva dicitur, et ceperunt eam"; Ralph de Diceto, *Ymagines Historiarum,* in his *Opera Historica* (ed. William Stubbs, London, 1876), II, 65: "Circa dies istos a partibus aquilonis naves plurimae sulcantes mare Britannicum foedus inierunt cum Anglis quos apud Dertesmu reppererunt. Itaque communi consilio XV° kalendas Junii se pelago commiserunt, cum essent naves XXXVII oneriferae multum et hominum multorum capaces. Postque varios rerum eventus III° kalendas Julii venerunt Ulixibonam. Rex autem Portugalensis . . . supplicavit ut ei venirent in auxilium ad subjugandum civitatem quandam Silviam."

[18] MS. *utensibus.*

[19] May 19.

Sanduuic procedentes, Wuichesse [20] venimus; sed inde propter contrarietatem ventorum vix quarto die [21] Ernemithie [22] venimus. Sequenti die velificantes, media nocte [23] ad portum Deramithie [24] venimus. Ibi inventis quibusdam sociis, mane [25] dimissa Anglia versus Britanniam processimus; sed deficiente vento et quandoque in contrarium flante, sex diebus in mari fluctuavimus; sed sexto die [26] zefirus tempestuosus nostro itineri contrarius ad insulam modicam a pauperibus Britannis inhabitatam velificare compulit, que a Gallis Belile, a Britonibus Wechele dicitur.[27] Infra iam dictos sex dies, preter sollempnia missarum, et extra portum Pente[4v.]costen [28] cum merore celebravimus.

Iuxta eandem insulam octo diebus fuimus, et nono die [29] carbasa ventis committentes [30] satis spirantibus, usque ad noctem processimus; sed ne incaute terram inpingeremus, velis depositis, cum terra [31] appareat, tota nocte naves vi ventorum agitate sunt. Postera [32] die [33] ad Rochiel [34] opulentissimum Pictavie opidum [35] applicuimus. Et notandum quod recto tramite pretermisso a Sancto Matheo,[36] qui locus est

[20] MS. reading very nearly certain, though Gazzera read *Vorychesse* and Chroust read *Voriichesse*, which they identified with Porchester. A more probable identification would seem to be with Winchelsea, Sussex, one of the Cinque Ports.

[21] May 23.

[22] Yarmouth on the Solent near the western extremity of the Isle of Wight.

[23] May 24–25

[24] Dartmouth, Devonshire, long a customary port of departure for crusaders and pilgrims to Spain, Portugal, and the East. Cf. *De Expugnatione Lyxbonensi*, ed. C. W. David (New York, 1936), p. 52, note 3.

[25] Apparently May 25, though one may well wonder that it was possible to make the junction and proceed so promptly. Compare the passage from Ralph de Diceto quoted above, note 17. Ralph gives May 18th as the date of departure and the number of ships as thirty-seven; but he is evidently speaking of a group of vessels which departed from Dartmouth somewhat in advance of those with which the present narrative is concerned. They seem to have been overtaken at Lisbon. See below, p. 616, line 8.

[26] May 30.

[27] Belle Ile, department of Morbihan, outside Quiberon Bay, off the coast of Brittany. *Wechele* is obviously derived from the ancient name which appears as *Vindilis* or *Vindelis* in the *Itinerarium Antoninum* (*Itineraria Romana*, ed. Otto Cuntz, Leipzig, 1929, p. 81) and which is evidently of pre-Celtic origin. The common form in the eleventh and twelfth centuries was *Guedel*, but *Guezel* seems to appear in some manuscripts. See *Cartulaire de l'Abbaye de Sainte-Croix de Quimperlé* (ed. Léon Maitre and Paul de Berthou, 2nd ed., Paris and Rennes, 1904), pp. 102, 131, 287, 294, 299, 304; *Cartulaire de l'Abbaye de Redon* (ed. Aurélien de Courson, Paris, 1863), pp. 246, 334; cf. Louis Rosenzweig, *Dictionnaire Topographique du Département de Morbihan* (Paris, 1870), p. 9. For a philological discussion of the word see Kurth, p. 179, note 4.

[28] May 28.

[29] June 7.

[30] MS. *commitentes*.

[31] MS. *terra ut*; it seems necessary to suppress *ut*, as Chroust has done, as being impossible to construe.

[32] MS. *Postea.*

[33] June 8.

[34] La Rochelle.

[35] The word is regularly spelled with a single *p* throughout the manuscript.

[36] Pointe-Saint-Mathieu, department of Finistère, some ten or twelve miles from Brest.

quedam Britannie extremitas in mare producta, propter iniuriam ventorum sinuosas quasdam maris ambages peragravimus, tum etiam ut duces vie Rochiel conduceremus.[37] Sciendum etiam quod Britannia quam in duobus lateribus circuivimus novem habet episcopatus,[38] quorum soli tres lingua [39] utuntur Britannica,[40] nulli alii genti communi. Reliqui vero Gallorum ydioma participant.[41] Britannia [42] in regno Francorum est, conterminas habens Andagaviam et Pictaviam.

Uno autem die Rochiel manentes, sequenti aurora,[43] velis expassis, pelagus aggressi sumus; sed, ventis dissentientibus et in diversa nos rapientibus, novem dies in alto fluctuantes exegimus. Nec obmittendum quod una nocte fulminibus et tonitruis terribili in summitate mali nostri plures de sociis duas candelas per longam moram ardere viderunt.[44] Illud etiam adiciendum est, quod innumerabilis multitudo piscium equalium rumbis sex vel VII. pedum mira velocitate sepissime totis corporibus apparentes nostras naves transsierunt.[45] Nono die [46] portum intravimus,[47] prope quem castrum est regis Galicie Gozeun [48] et opidum

[37] Chroust has emended the clause to read *dum etiam per duo duces vie Rochiel conduceremur*, but this seems unnecessary and a violent alteration of the probable meaning of the text. The mention of pilots (*duces vie*), if pilots they be, is of much interest and should be compared with the reference to a Gallegan *dux cuiusdam navis nostre* on p. 624, line 9, below.

[38] The nine dioceses of Nantes, Rennes, Dol, Saint-Malo ·Saint-Brieuc, Tréguier , Léon (Saint-Pol), Cornouaille (Quimper), and Vannes. Brittany was in effect an ecclesiastical province under the archiepiscopal jurisdiction of Dol from the middle of the ninth to the end of the twelfth century, though the fact was never formally recognized by the papacy and was disputed by the archbishops of Tours and other anti-Breton interests, and in 1199 Pope Innocent III formally suppressed the claims of Dol and declared all the bishops of Brittany subject to the archbishop of Tours. See A. de la Borderie, *Histoire de Bretagne* (Rennes, etc., 1896–1914), III, pp. 197–205, and the convenient map *ibid.*, vol. I, at the end.

[39] MS. *ligua.*

[40] MS. *Britannia.*

[41] The three Breton-speaking dioceses evidently were Tréguier, Léon, and Cornouaille, though the evidence of modern speech would seem to indicate that a good part of the diocese of Vannes was also included in the region to which Breton speech was confined as a result of the Norse invasions in the ninth and tenth centuries. On the linguistic frontier see J. Loth, *Chrestomathie Bretonne* (Paris, 1890), p. 237; La Borderie, *op. cit.*, III, pp. 215–217; and on the Norse invasions, *ibid.*, II, pp. 299–373.

[42] MS. *Britanniam.*

[43] June 9.

[44] An example of St. Elmo's fire.

[45] The fish in question were doubtless dolphins or porpoises, which in their general appearance as observed from shipboard might well be thought to resemble sturgeons (*rumbi*). For the true definition of *rumbus* as sturgeon see K. Höhlbaum, *Hansisches Urkundenbuch*, III, (1882–86), Glossary, to which reference is made by Allan Evans in his edition of Francesco Balducci Pegolotti, *La Pratica della Mercatura* (Cambridge, Mass., 1936), p. 253, note 9.

[46] June 18.

[47] The port of Luanco, province of Oviedo, which in the Middle Ages was known as Gozón, or Gauzon, from the near-by castle of that name. The name Gozón now survives only as that of the municipal district (*municipio*) lying between Cape Peñas, Avilés, and Luanco. The last, which is the administrative centre of the district, is situated at the head of Luanco Bay which lies between Cabrito and Castillo Points, on the latter of which are the ruins of a castle. A direct road leads from Luanco to Avilés which is about eight miles away. See U. S. Hydrographic Office, *Bay of Biscay Pilot* (3rd ed., Washington, 1926), p. 484, and U. S. Hydrographic

Abilez.[49] Et notandum quod predictis novem diebus Waschoni[5r.]am, regnum Aragon*ens*ium,[50] regnum Navarrorum, regnum Ispanie a sinistris reliquimus, et iam in regno Galicie fuimus. Considerandum etiam quod, cum sint quinque regna Ispaniorum,[51] videlicet Arragonen*sium*,[52] Navarrorum et eorum qui specificato vocabulo Ispani dicuntur, quorum metropolis est Tolletum,[53] item incholarum Galicie et Portugalensium, et ea [54] ex omni latere preterquam ex uno ambiat mare, omnia habent terminos versus mare Britannicum, per quod venimus, et limites habent contra Sarracenos, qui [habitant] [55] in margine oppositi maris; et ita qui ad ultimum, id est Portugalensium regnum, ire vellet, per omnia transsire deberet.

Decimo die [56] naves in portu relinquentes, ad Sanctum Salvatorem [57] profecti sumus, civitatem que a portu sex leucis distat.[58] Ibidem invenimus archam reppletam diversis magna veneratione dignis et sanctorum reliquiis que tempore persecutionis [ ] [59] propter metum

Office charts, nos. 4379 and 4380; *Enciclopedia Universal Ilustrada Europeo-Americana* (Barcelona: J. Espasa, 1912–1929), XXV, 1094, XXVI, 844; *De Expugnatione Lyxbonensi*, ed. David, pp. 60–61, note 3.

[48] Chroust mistakenly read *Goyeun* and identified the place with Gijón; but the correct identification is with the castle of Gozón, or Gauzon, which may perhaps be identified with the ruins on Castillo Point on the north side of Luanco Bay; cf. note 47 above and the references there cited.

[49] Avilés, province of Oviedo, on the Avilés River some two or three miles from its mouth; cf. note 47 above.

[50] The extension of the abbreviation is uncertain; Chroust read *Aragonium* and emended the text to read *Aragonum*.

[51] "Los cinco reinos de España." Cf. R. Menéndez Pidal, *La España del Cid* (Madrid, 1929), II, 687–689.

[52] The extension of the abbreviation is uncertain; Chroust read *Arragonense*.

[53] Toledo.

[54] MS. *eos*.

[55] I follow Gazzera in supplying *habitant* which seems wanting to complete the sense.

[56] June 19.

[57] Oviedo, see below, p. 615, lines 2–3.

[58] Actually somewhat more than twenty-five miles by the modern road.

[59] The manuscript at this point is unintelligible: *persecutionis* is clear, the last three letters being written at the beginning of a new line, and a space of about one centimetre being left after them, apparently by the original scribe. In this space there has been inserted, in conspicuously different ink, presumably by an ununderstanding corrector, something which, taken with *nis* of *persecutionis*, may have been intended to make *misericordie*. Evidently the writer, either the original author working with some account of the Oviedo relics before him, or a later copyist, had to do with something which was unintelligible to him. Possibly the word *gentilium* or the word *Mahometi* should be supplied after *persecutionis;* cf. the various accounts of the relics indicated by the following citations: a letter of Osmond, bishop of Astorga (1082–1096) to Ida, countess of Boulogne, quoted by Ch. Kohler in *Revue de l'Orient Latin*, V (1897), 3–4; *Historia Silense* (ed. F. Santos Coco, Madrid, 1921), pp. 23–24; Lucas of Tuy, *Chronicon Mundi*, in Andreas Schott, *Hispania Illustrata* (Frankfort, 1603–08), IV, 74; Rodrigo Jiménez de Rada, *De Rebus Hispaniae*, in *Rerum Hispanicarum Scriptores . . . ex Bibliotheca Roberti Beli* (Frankfort, 1579–80), I, 203; *Primera Crónica General* (ed. R. Menéndez Pidal, Madrid, 1906), I, 348. Gazzera passed the difficulty with the misleading notation "Vox oliterata"; Chroust read *eodem* after *persecutionis*, but, as it seems to me, erroneously.

hostilem ab Iherosolima translata [sunt] [60] in Affricam, inde in Ispalim, que nunc Sibilia,[61] ab Isspali in Tolletum, a Tolleto in Ovetum,[62] quod nunc Sancti Salvatoris nomine pretitulatur.[63] Nota quod in costa Galicie non nisi arduas rupes vidimus,[64] nam et ipsa tota valde montuosa est, et ideo sterilis,[65] et non vinifera, cicera maxime utens.

Undecimo die [66] ad naves regressi et tercio decimo [67] circa crepusculum ad mare reversi sumus. Quarto decimo, id est in vigilia Iohannis Babtiste,[68] et [69] ipso festo valido flatu velis turgentibus in vespera diei sancti [70] ad portum venimus [71] Tambre,[72] que est aqua fluens per Galiciam. Ibi relictis navibus, per longam dietam regressi, limina sancti [73] Iacobi, [5v.] que iam transsieramus, visitavimus.[74] In eundo autem et redeundo et moram in portu pro ventorum prestolatione faciendo [VIII] dies [75] peregimus.

In octava Iohannis [76] circa meridiem naves conscendimus, et sequenti meridie [77] Portugaliam [78] e vicino vidimus. Inde ventis prospere spirantibus tercio [79] die diluculo portum Ulixibone [80] intravimus, qui portus est ostium [81] Tagi,[82] a Tholeto venientis et in mare

[60] I have supplied *sunt* which the sense seems to require.

[61] Seville.

[62] Oviedo.

[63] On the early history of these famous relics see, besides the references cited above, note 59, Enrique Flórez, *España Sagrada*, (Madrid, 1749–1879), XXXVII 279–294; on their recent vicissitudes see *De Expugnatione Lyxbonensi*, ed. David, p. 62, note 2.

[64] On the character of this coast see U. S. Hydrographic Office, *Bay of Biscay Pilot*, p. 465, and cf. *De Expugnatione Lyxbonensi*, ed. David, pp. 60–61, note 3.

[65] Chroust accepted the erroneous reading of Gazzera, *difficilis*, but the correct reading is clear in the manuscript.

[66] June 20.

[67] June 22.

[68] June 23.

[69] Chroust supplied *in* after *et*.

[70] June 24.

[71] MS. *invenimus*.

[72] Muros or Noya Bay at the mouth of the Tambre River on the west coast of northwestern Spain, some miles to the south of Cape Finisterre.

[73] MS. apparently *sibi*, by a scribal error.

[74] Santiago de Compostela is some twenty-five miles eastward and slightly north of the mouth of the Tambre.

[75] MS. *radies*, written with perfect clearness, which makes no sense. Evidently the copyist had before him the word *dies* preceded by a numeral which he wholly misunderstood. Since the voyage was continued on the octave of St. John, it may be inferred that the numeral was *VIII*, and this I have supplied. Chroust, following Gazzera, supplied *VII*, remarking in a note that the number was not decipherable.

[76] July 1.

[77] July 2.

[78] Oporto.

[79] July 4 if the author is reckoning from the 2nd when they were off Oporto, July 3 if he is reckoning from the 1st on which they sailed from the mouth of the Tambre.

[80] Lisbon.

[81] MS. *hostium*.

[82] The Tagus River.

ibi fluentis. Est autem amplum sicut Albia [83] iuxta [84] Stadium.[85] Nota iuxta Ulixibonam ad tria nostra miliaria est castrum nomine Sintricum,[86] ubi concipiunt eque de vento,[87] et sunt fetus velocissimi sed non ultra octo annos viventes. Hec Ulixibona, opulenta et magna valde, ante quadraginta et $\overset{\text{or}}{\text{IIII}}$ annos a peregrinis nostris capta [88] cum adiacentibus castris subiacet dominio [regis] [89] Portugalensis. Terra illa satis fertilis et sana est, apte montibus erecte et satis in valles protensa.[90]

Ibi invenimus naves $\overset{\text{or}}{\text{XXIIII}}$.,[91] et nos undecim [92] habuimus. Sed processerant nos ante IIII. ebdomadas, vel V.,[93] naves de nostro imperio et de Flandria; et in [94] itinere ultra Ulixibonam castrum quod subiacebat dominio Silvie, Alvor [95] nomine, expugnaverunt, nulli etati vel sexui [96] parcentes; et, sicut veraciter audivimus, circiter V. milia et

[83] The Elbe River.

[84] MS. *iusta.*

[85] Stade, Prussian province of Hanover, is actually on the Schwinge, three and one-half miles above its confluence with the Elbe at a point some twenty miles below Hamburg. It was formerly the chief port of Hanover.

[86] Cintra is actually about seventeen miles from Lisbon. By *nostra miliaria* the author evidently meant the long German miles with which he was familiar at home. On p. 618, line 1, below he uses the phrase *miliare Teutonicum.* This German mile may have equaled about 7,500 metres, or somewhat more than four English statute miles; cf. *Meyers Konversations-Lexicon* (6th ed., Leipzig and Vienna, 1902–09), *s. v. Meile.*

[87] On this curious legend see *De Expugnatione Lyxbonensi*, ed. David, p. 93, note 6.

[88] The reference is to the conquest of Lisbon in 1147 (actually forty-two years earlier) on which see *De Expugnatione Lyxbonensi*, ed. David. Kurth (pp. 164–165) interprets the clause as meaning forty-four years before the time when the author was writing, thereby indicating 1191 as the composition date of the work; but Chroust (p. ci) suggests that the author has simply made a mistake, a hypothesis which seems much more likely.

[89] I have supplied *regis* which the sense seems to require. Chroust met the difficulty by emending *Portugalensis* to read *Portugalensi.*

[90] MS. *procensa;* there is no warrant for *pretensa*, the reading of Gazzera.

[91] These may have been some of the thirty-seven ships which, according to Ralph de Diceto, sailed from Dartmouth on May 18 and reached Lisbon on June 29; see the passage quoted above, p. 611, note 17.

[92] These were not the identical eleven vessels which sailed originally from Blexen. Two of those had been lost at Sandwich and only one of these had been replaced through an acquisition made in London. Thus it seems that the squadron must have consisted of ten vessels as it sailed from Dartmouth for Spain. The eleventh vessel is perhaps to be accounted for by the galley of Tuy referred to in the text below, p. 617, line 7.

[93] The reading of the manuscript is clear and certain, although Chroust, following Gazzera, read *quinquaginta quinque* in place of *vel V.* He was perhaps influenced thereto by a reference to a fleet of fifty-five vessels in the *Annales* of Lambertus Parvus which is referred to by Kurth (p. 172, note 1); it is, of course, possible that the copyist blundered and that the text originally read *LV*, but of this there is no evidence. The Turin manuscript as it stands gives no indication of the number of ships participating in the conquest of Alvor. For the conquest of Alvor see below, Appendix B.

[94] MS. *in* inserted above the line in different ink.

[95] Alvor, the ancient *Portus Hannibalis*, at the mouth of the River Alvor, on the south coast of Portugal some five miles northeast of Lagos and perhaps twenty-five miles from Cape St. Vincent; see map facing p. 666; cf. Biblioteca Nacional de Lisboa, *Guia de Portugal* (Lisbon, 1924–), II, 280; F. X. d'A. Oliveira, *A Monografia de Alvor* (Oporto, 1907).

[96] MS. *sexsui.*

sexcentos occiderunt.[97] Galee autem de Ulixibona eas comitate usque ad strictum mare,[98] tandem reverse, ipsas prospere procedere nobis renunciaverunt, et aliquos Serracenos captivos reduxerunt. Nos interea accingebamur ad obsidionem [99] Silvie pro petitione [regis] [100] Portugalensis [101] cum multis copiis etiam se properantis.[102] Morati [103] autem [sumus] in portu XI. diebus cum XXXVI.[104] navibus magnis et una galea de Tue,[105] opido Galicie, que nobis contubernio [106] iuncta est, et pluribus navibus de Ulixibona.[107] Circa vesperam undecime diei [108] profecti, tribus diebus et duabus noctibus [109] continue, sed lente, quando [110] velificavimus, tercia die [111] post meridiem vidimus Alvor castrum, quod nostri expugnaverant, destructum, supra mare situm, et alia quedam deserta quorum incole in Alvor occisi erant. Non longe inde intravimus portum Silvie,[112] terram optime cultam invenientes, sed habitatores omnes confugerant in Silviam. Distat autem Silvia a

[97] See below, Appendix B.

[98] The Strait of Gibraltar.

[99] MS. *accingebantur ab obsidione*. I adopt the emendation of Gazzera and Chroust.

[100] I follow Gazzera in supplying *regis*, though Chroust did not.

[101] Sancho I, king of Portugal 1185–1211, son and successor of Alfonso Henriques, the founder of the monarchy.

[102] Chroust, perhaps rightly, emended to read *preparantis*. For a reference to the terms of the original agreement between Sancho and the crusaders, which did not remain unaltered to the end of the siege of Silves, see below, p. 631, lines 6–7. Ralph de Diceto (*Opera Historica*, ed. Stubbs, II, 65–66) gives the agreement as follows: "[Rex] pactum iniit cum eis tam ipse quam tres episcopi sui, praestito sacramento, quod quicquid vel auri vel argenti vel victualium civitate subacta possent adquirere, suos in usus redigerent, et solam urbem regis potestati subigerent." Cf. *El Anónimo de Madrid y Copenhague*, ed. A. Huici, p. 61; 'Abd al-Wāhid al-Marrākushī, *Histoire des Almohades* (French translation by E. Fagnan), p. 243.

[103] MS. *moranti;* I adopt the emendation of Gazzera and Chroust and supply *sumus* after *autem* which the sense seems to require.

[104] MS. *XXVI* was written first and then changed to XXXVI, the correction being made with darker ink.

[105] Tuy, Spanish province of Pontevedra, on the right bank of the Minho, directly across the river from Portugal. The reading of the manuscript is very nearly certain, and Chroust read the word correctly though he failed to make any modern identification, apparently being confused by the speculations of his predecessors. Gazzera unfortunately read *Rue* and was thereby misled as to the true identification, as were also Silva Lopes and Kurth (p. 183, note 2).

[106] MS. *conturbernio.*

[107] According to Ralph de Diceto (*Opera Historica*, ed. Stubbs, II, 65) the king contributed thirty-seven galleys and a considerable number of *sagittariae* to the expedition. The author of the present text confirms the presence of Portuguese galleys and *sagittariae* (which he calls *sagiccinae*, for *sagittinae*) at the siege of Silves, but he is evidently reluctant to admit the size and importance of the Portuguese forces. See below, p. 618, line 8, p. 619, line 9, p. 630, line 8.

[108] July 14.

[109] MS. possibly *tribus diebus et duabus noctibus* has inadvertently been written for *tribus noctibus et duabus diebus.*

[110] Chroust emended *quando* to read *quidem* and continued the previous sentence through *velificavimus.*

[111] July 17.

[112] July 17; the date is confirmed by Ralph de Diceto, *Opera Historica*, ed. Stubbs, II, 66.

mari via terrestri per miliare Teutonicum,[113] sed tortuosior et ideo longior est via in aqua.

Nostri ergo per inimicorum terras nimis avide et incaute discurrerunt,[114] et ideo duo Bremenses,[115] ab aliis stulte [116] separati, a decem equitibus Sarracenis, [6r.]quos solos in tota terra vidimus, occisi sunt. Sed ad classem reportati, ibidem a nobis sepulti sunt. Nostri ergo in portu non longe a mari anchoris fixis, villas exusserunt, et pauca que reperierunt diripiebant.[117] Nocte [118] autem sagiccinam [119] unam de Ulixibona pro principe milicie Portugalensis [120] misimus, qui per terram profectus [121] nos precesserat,[122] et tunc castra sua distabant a nobis per III (or IIII) miliaria.[123] Sequenti die [124] navis una peregrinorum de Britannia venit ad nos.[125] Eodem die princeps [126] milicie Portugalensis [127] circa

[113] On the German mile see above, p. 616, note 86. Silves is actually about six and a half miles from the coast by the modern road; it is between seven and eight miles by the river.

[114] MS. *discarrerunt.*

[115] MS. reading clear; there is no warrant for *Brenienses*, the reading of Gazzera which was accepted with a query by Chroust who, however, adopted *Bremenses* as the correct form of the text. The presence of men from Bremen is, of course, natural, in view of the facts related above, p. 610, line 12.

[116] MS. *stulti.*

[117] For detailed maps showing the situation see British Admiralty charts, No. 2680; U. S. Hydrographic Office charts, No. 4402; Instituto Geografico e Cadastral, *Carta de Portugal*, scale 1:50,000, sections 49C, 49D, 52A, 52B (new numbering). The map facing this page is based on the latter.

[118] July 17.

[119] Evidently the same as *sagittina* (diminutive of *sagitta*, Italian *saettia*) on which see A. Jal, *Glossaire Nautique* (Paris, 1848). Presumably the *sagittariae* referred to by Ralph de Diceto, above, p. 617, note 107, were vessels of the same kind.

[120] It does not seem possible to determine with any certainty who this Portuguese commander was; cf. Herculano, *Historia de Portugal*, 8th ed., III, 348–352; Kurth, p. 184, note 2. The reference to "Peter, son of Henry" by 'Abd al-Waḥīd al-Marrākushī (French translation by E. Fagnan, p. 243) which impressed Herculano seems to the present editor to be probably due to a misunderstanding, or possibly to a corrupt text.

[121] Cf. below, p. 635, lines 1–5, where the author describes the land route from Lisbon to Silves as a journey of seven days, through country in which there was not a safe habitation for Christian or Saracen because of the danger of raids by one people against the other. Idrīsī gives the distance as six days' journey. See below, p. 635, note 345.

[122] MS. *processerat.* I adopt the emendation of Gazzera and Chroust.

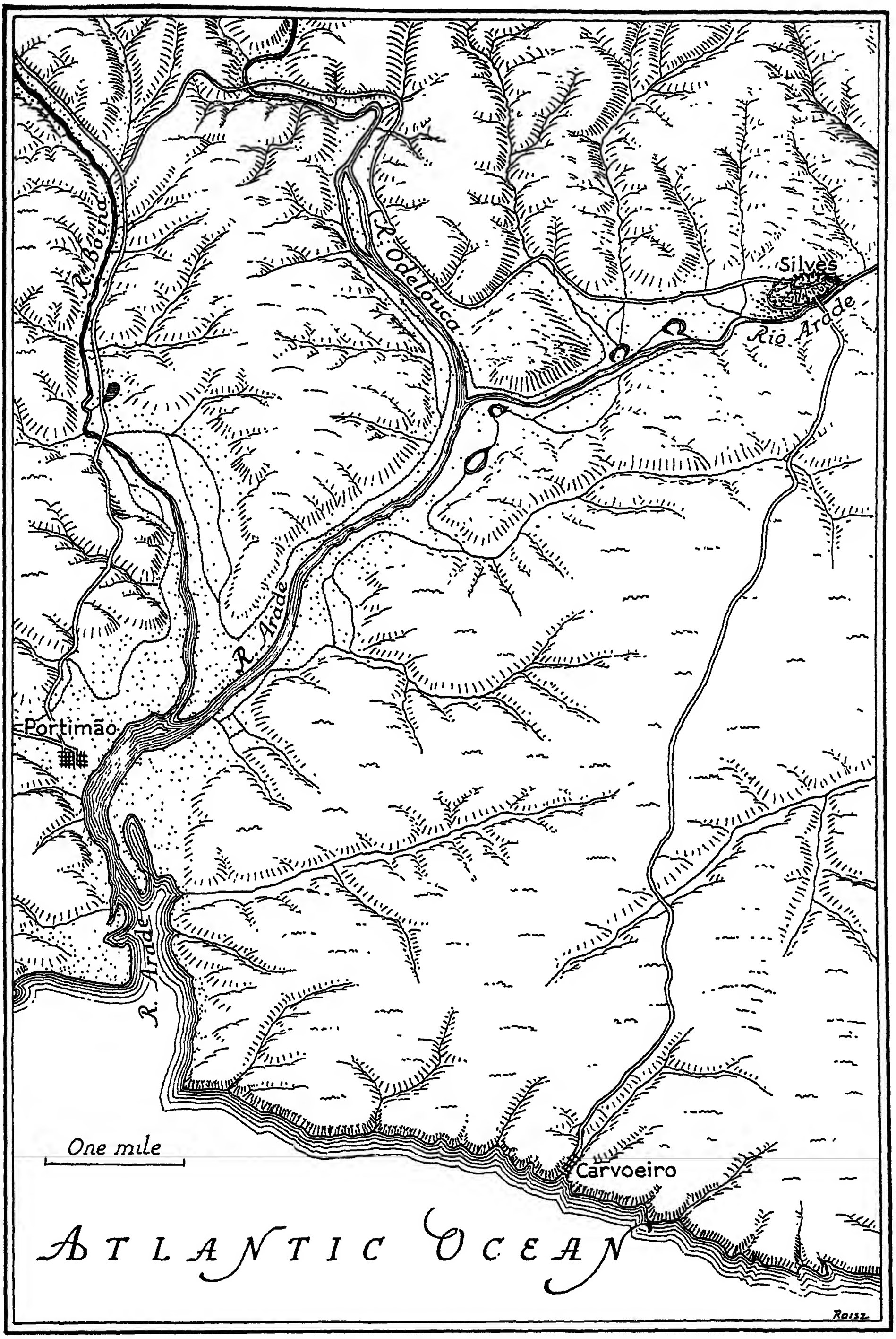
[123] Presumably the long German mile is meant. It is tempting to conjecture that the Portuguese commander had reached a point on the Odelouca River at the head of navigation, but this is a mere guess. The Odelouca is navigable for small craft as far as the modern bridge from which a direct road leads to Silves. See U. S. Hydrographic Office, *East Atlantic Pilot* (4th ed., Washington, 1929), p. 177. See map facing this page.

[124] July 18.

[125] With the arrival of this ship from Brittany the besieging naval force was, so far as there is any evidence to show, complete. Including this last ship to arrive, the northern contingent would seem to have consisted of thirty-seven vessels, to which there is to be added the galley of Tuy which was closely associated with the northerners, making thirty-eight in all. The Portuguese naval forces were also considerable. According to Ralph de Diceto they consisted of thirty-seven galleys and a considerable number of *sagittariae;* see above, p. 617, note 107.

[126] MS. *principes.*

[127] MS. a line drawn through this word, perhaps for the purpose of deletion.



Environs of Silves to Illustrate the Siege.

vesperam cum paucis venit, relicta milicia in castris. Cum ergo discuteretur quid agendum esset, magis placuit ei ut ad Dardeam [128] capiendam proficisceremur,[129] nam desperabat a nobis posse capi Silviam, utpote illius regni metropolim [130] et valde firmam. Nos autem magis confisi in Domino fiducialiter, maioris opere laborem aggressi sumus, quod et princeps [131] ratum habuit. Postera [132] die [133] cum navibus ad civitatem processimus, et anchoras [134] fiximus in loco unde eam videre potuimus,[135] sed procedere nos [136] negabat aque raritas. Princeps etiam milicie ultra nos excubabat cum suis et cum galeis que nobiscum venerant.[137] Illa nocte [138] accenderunt multa luminaria in civitate, et nos similiter; et letissimus erat populus noster, non formidans quod munitissimum vidit locum. Ulteriori die [139] diluculo, armati cum scaphis, ad civitatem accessimus et castra fiximus ita ut de muro facile bis [140] cum balista iaceretur ad castra.[141]

Silvie autem status talis est.[142] Magnitudine non multum dissidet a Goslaria,[143] sed multo plures domos habuit [144] et amoenissimas [145] manssiones, muris cincta et fossatis, ita ut solum tugurium extra muros non inveniretur. Et quatuor erant intus munitiones, quarum prima civitas ampla in valle quam [146] rovalle [147] dicunt; civitas in monte quam

[128] All efforts to identify this place have proved fruitless. There is no warrant in the manuscript for *Gardeam*, the reading of Gazzera.

[129] MS. *proficeremur*.

[130] Silves at this period would have been more properly described as the metropolis of a province of the Almohade empire, the province of al-Gharb, though it had been for a brief period in the middle of the eleventh century the capital of an independent petty kingdom. See Appendix A.

[131] MS. *principes*.

[132] MS. *postea*.

[133] July 19.

[134] MS. *anchoris*.

[135] Silves becomes visible between hills a short distance below the confluence of the Odelouca with the Arade. See map facing p. 618.

[136] MS. *non*.

[137] Apparently the galleys, drawing less water, had been able to pass the shallows and join the Portuguese forces somewhat farther up stream.

[138] July 19–20.

[139] July 20.

[140] MS. *bis* written after *cum balista;* signs placed above the line indicate its proper position in the text. Gazzera read *lapis* without warrant.

[141] Presumably the crusaders' camp was pitched on the flat land on the right bank of the Arade some distance below the wall of the western suburb which the author in the next paragraph calls *prima civitas* or *rovalle*.

[142] See Appendix A on the Situation and Mediaeval History of Silves.

[143] Goslar, Prussian province of Hanover.

[144] Chroust has emended to read *habet*.

[145] MS. *hamenissimas*.

[146] MS. *qua*.

[147] Evidently derived from Arabic *al-rabaḍ* (Portuguese *arrabalde*, Spanish *arrabal*), meaning suburb; cf. E. Lévi-Provençal, *L'Espagne Musulmane au Xème Siècle* (Paris, 1932), p. 151; *idem*, *Inscriptions Arabes d'Espagne* (Leyden and Paris, 1931), p. 52. This suburb was doubtless located to the west and southwest of the hill and the main fortifications of the city,

almadinam[148] vocant, habens aliam munitionem proclivem[149] in rovalle, descendentem ad conductus aquarum[150] et cuiusdam fluvii qui dicitur Widradi[151] (alius in eundem fluit qui dicitur Wideloc);[152] et super conductum $\overset{\text{or}}{\text{IIII}}$ turres ita ut superior civitas inde aquis habundaret, et hec munitio corrasce[153] dicitur. Introitus portarum ita angulosi et tortuosi erant ut facilius transsiretur murus quam ostium[154] intraretur. Item sub primum castrum alcaz[155] dicebatur.[156] Item

on the broad level ground that spreads out from the right bank of the river. It must have been entirely outside any surviving fortifications and doubtless included the present landing place for small boats which still come up the river at high tide.

[148] MS. *almadina.* Derived from Arabic *al-madīnah,* meaning city and apparently used in the Iberian peninsula in this period to designate the main fortification of a city.

[149] A fortification descending the southern slope of the steep hill from the main fortress (*almadina*) to the four towers (*corrasce*) which protected the water supply (*conductus aquarum*) mentioned below.

[150] The water conduit which supplied the city with fresh water and is perhaps to be identified with the present mill-race which supplies power to the Moinho da Porta just above the ancient bridge over the Arade. This is the plausible view expressed by Pedro P. Mascarenhas Júdice in his recent work *A Sé e o Castelo de Silves* (Gaia, Portugal, 1934), pp. 58–61, quoting an earlier article which he published in the *Voz do Sul* of Silves, December 23, 1929. His argument briefly is as follows: The twelfth-century geographer Idrīsī in his description of Silves says: "On y boit l'eau d'une rivière qui vient à la ville du coté du midi et qui fait tourner des moulins." Idrīsī's reference cannot be to the Arade itself at Silves, since the tide comes right up to the rocky barrier above the bridge, the so-called dam (*chamada comporta*), and renders the water undrinkable. The present mill-race which branches off from the right bank of the Arade some distance above the town and flows down parallel with the river, bringing fresh water and driving the mill (as it may have done in the twelfth century), seems to offer an adequate and harmonious explanation of Idrīsī's text and of the phrase *conductus aquarum* as it figures in the present text here and below.

On the other hand, it is to be noted that on p. 625, line 9, below, this same source of water supply is described as a *puteus* and that Ralph de Diceto (*Opera Historica,* ed. Stubbs, II, 66), who was not an eye-witness, describes it as follows: "fontem . . . duplici muro circumdatum, habentem barbacanam novem turribus circumseptam."

[151] The Arade River, sometimes called the Portemão River, and sometimes also the Silves River.

[152] The Odelouca River which flows into the Arade several miles below Silves.

[153] Evidently related to Portuguese *couraça* or *coiraça,* a defensive work thrown up around a besieged city, a breastwork. No part of this construction seems to have survived.

[154] MS. *hostium.*

[155] Derived from Arabic *al-ḳaṣr* (Portuguese *alcaçar,* Spanish *alcázar*), meaning a fortress.

[156] This sentence has offered serious difficulties to editors and users of the text. Gazzera printed it as it stands in the manuscript, except for a misreading of *alcaz* as *alcay.* Herculano (*Historia de Portugal,* 8th ed., III, 172, note 2), feeling that the text must be corrupt, proposed to emend it to read *Super primam* (*scil. civitatem*) *castrum Alcay dicebatur;* and this proposal was in the main accepted by Kurth (p. 189, note 3), with the result that Chroust was influenced to print the text as follows: *Item* [*munitio*] *super primam* [*civitatem*] *castrum alcaz dicebatur.* That such a rendering is impossible is proved by the use of the words *prima civitas* higher up to describe the suburb (*rovalle*) rather than the main fortress on the hilltop (*almadina*) to which Chroust evidently would refer. When the author wishes to refer to this upper fortress he calls it "the upper city" (*superior civitas*).

The text as it stands in the manuscript may indeed be corrupt, but it is by no means certainly so. While many of the surviving towers of Silves measure about sixteen feet square, there is one great tower, measuring nearly forty feet square, which stands beneath what is to this day called "the port" and was evidently the original main entrance (*introitus*) to the principal fortress (*almadina*). If this great tower was the *alcaçar* or if the author supposed it to be such, the passage offers no particular difficulty.

una magna turris [157] erat in rovalle, et habebat viam ad almadinam, id est [158] muro quodam testudinato,[159] ita ut de ea videri posset quid exterius muro almadine accideret, et impugnantes [160] murum a tergo ledi possent a turri, et e converso; [161] et hec alve[6v.]rana [162] dicitur. (Et nota quod hec nomina sunt appellativa, non propria. Ubicumque enim tales dispositiones sunt civitatis in terra illa, tam a Christianis quam a paganis,[163] talia habent nomina. Sarraceni autem in Ispania habitantes Andeluci [164] dicuntur, qui in Affrica, Mucimiti vel Maximiti vel Moedini,[165] qui in Marorlce,[166] Moravidi [167].) Nota etiam quod in muro cuiuslibet munitionis erant turres ita vicine ut modicus fuerit iactus lapidis de una turri ad terciam. In quibusdam locis duplo viciniores erant turres.

Illa ergo die [168] statim cum veneramus, egressi de civitate circiter decem equites, discurrebant prope murum, quasi nostros provocantes; unde minus providi propter iussionem magistrorum ad ipsos currebant; sed a muro telis et lapidibus arcebantur, vulnerabant autem et vulnerabantur, et dubio Marte revertebantur. Nos autem propius figentes temptoria [169] consilium inivimus, ut mane insultum faceremus, preparantes scalas ad transseundum murum. Diluculo [170] igitur celebratis missarum [171] sollempniis, cum populus cum multa devotione communicasset,[172] armatus cum scalis ad murum accessit; et ultra fossatum

[157] All traces of this watch-tower (*alverrana*) seem to have disappeared, as have also all traces of the walls of the suburb (*rovalle*) itself.

[158] MS. reading *id est* not perfectly certain but very probable; Chroust, following Gazzera, read *in*.

[159] MS. *testudinoto*.

[160] MS. *impugnates*.

[161] MS. *et everso;* Chroust emended to read *e converso*, omitting *et*.

[162] Derived from Arabic *al-barrānīya* (Portuguese and Spanish *albarrana*), meaning an external tower, i.e., a tower situated outside a wall or fortification; hence a watch tower. Cf. E. Lévi-Provençal, *L'Espagne Musulmane au Xème Siècle*, p. 150, note 6.

[163] MS. *ragones;* Chroust emended to read *agarenis*.

[164] Andalusians.

[165] *Mucimiti* and *Maximiti* are evidently corruptions of *Masmuda* or *Maṣmūdah*, the name of the Berber tribe from which the founder of the Almohade movement was sprung and by which the Almohades (*al-Muwaḥḥidūn*) were not infrequently called; cf. R. Dozy, *Recherches sur l'Histoire et la Littérature de l'Espagne pendant le Moyen Âge* (3rd ed., Paris and Leyden, 1881), II, 438 and note 3; 447, note 5; Stanley Lane-Poole, *The Mohammadan Dynasties* (reprint, Paris, 1925), p. 45; P. K. Hitti, *History of the Arabs* (London, 1937), p. 546. *Moedini*, on the other hand, doubtless stands for *Muwaḥḥidūn*, unitarians, the religious designation of the Almohades.

[166] Morocco.

[167] Almoravides.

[168] July 20.

[169] The tents were now moved up towards the wall of the suburb (*rovalle*) but not up to it. For the next move, after the suburb had been taken, see below, p. 623, lines 3–4.

[170] July 21.

[171] MS. *celebratissimarum*.

[172] MS. *communicasset* written after *scalis* and signs placed above the line to indicate the proper order.

reptans, nec profunditatem aquarum declinans, ad murum venit. Qui autem in propugnaculis [173] erant aliquamdiu saxa iacientes, repente tandem Dei nutu qui "salvat sperantes in se," [174] in fugam conversi ad superiorem locum confugerunt. Nostri autem ilico scalis appositis, eos sunt insequuti.[175] Sed quia maturius [176] fugerant, pauci [177] occisi, alii in munitionem [178] sunt recepti, tum etiam quia nostri armati, illi inermes [erant], [179] et ideo facilius evaserunt. Multi tamen in porta pre nimia festinatione suffocati sunt, quorum corpora extra muros proiecerunt, nescio quare, nolentes forsitan [180] sepelire. Deinde etiam [nuntiatum] [181] nobis erat quod rex suus [182] fecit decollari qui primo fugerunt.[183] Sic [184] ergo possedimus civitatem inferiorem, ex una parte nostri, ex alia parte Portugalenses; et illa die et nocte in civitate quievimus.

Sequenti aurora, id est in festo beate Marie Magdalene,[185] premissis missa et communione,[186] nostri armati egressi sunt civitatem, relictis intus galiotis; et scalas secum ferentes, insultum fecerunt in civitatem superiorem ubi firmissima erat, monti imposita et fossato profundo et precipiti cincta.[187] Nostri autem instanti et diuturno labore scalas apponere moliti, profunditate fossati [188] et creberrimis[7r.] iactibus repellebantur,[189] licet multi in propugnaculis [190] assiduis nostrorum sagittis vulnerarentur. Concepta igitur spe frustrati, inconsulcius quam deceret, districtis animis, captam civitatem, quantum potuimus, concremavimus; quia talis erat materia domorum quod cum arderet una domus, alia ideo non incenderetur; nam tectum latericium, parietes lutei, cemento vestiti, et pauci ligna habebant. Et V. galeas cum aliis

[173] MS. *pugnaculis;* I adopt the emendation of Chroust.

[174] Dan. 13 : 60.

[175] MS. *insequti.*

[176] MS. *muturius*, though the reading is doubtful; *multi* was apparently written first, and then imperfectly corrected by a partial erasure of *l* and by the addition of two marks of abbreviation. I adopt the emendation of Chroust.

[177] MS. *paci.*

[178] MS. *munitione.*

[179] I follow Chroust in supplying *erant* which the sense seems to require.

[180] MS. possibly *forsan,* as Chroust has read.

[181] I follow Gazzera and Chroust in supplying *nuntiatum*, though *auditum* would perhaps be a more likely word.

[182] Chroust has emended to read *eos.*

[183] MS. *fugierunt.*

[184] MS. *Si.*

[185] July 22.

[186] MS. *premissas missas et communionem.*

[187] Faint traces of a dry moat are still visible high on the hill underneath the northern wall of the main fortress. For a specific reference to this moat, on the north side, see below, p. 626, lines 11–12.

[188] MS. *fossate.*

[189] MS. *reppellebatur.*

[190] MS. *pugnaculis;* I adopt the emendation of Chroust.

navibus pre timore hostili infra muros receptas incendimus,[191] et ad priora castra reversi sumus.

Sed eodem die resumptis animis et bellandi constancia, castra iuxta murum capte urbis posuimus;[192] et pluribus diebus machinas ligneas, turres, scalas, et diversa instrumenta ad capiendam urbem instruximus. Interea etiam multiplicabatur exercitus Portugalensium[193] in obsidione[194] nobiscum excubantium.[195] In octava Marie Magdalene[196] venit Portugalensis rex, et exercitus suus cum sarcinis lente subsequutus[197] est. Proxima die,[198] que erat dominica, quia quidam ex nostris Anglici[199] occiderant Sarracenum ante biduum in oculis eorum qui obsessi erant, ipsi nobis videntibus de turri que alverrana dicitur,[200] suspenderunt tres Christianos per pedes, quos ante habuerant captivos, et gladiis et lanceis usque ad mortem percusserunt. Unde lacrimosa compassione doluimus et ad bellandum magis exasperati[201] sumus. His[202] diebus augebatur exercitus Portugalensium,[203] et circumdata est undique civitas obsidione; nec aliquando requievimus vel instrumentum facientes vel sagittis et machinis impugnati vel impugnantes.

Die dominico in festo Sixti, Felicissimi, et Agapiti,[204] nos de regno Teutonico diluculo impulimus quoddam instrumentum, quod ericium vocamus, ad muros corrasce inter duas turres, ut perfoderemus murum. Instrumentum autem illud firmissimum erat, magnis lignis compactum, et novis gubernaculis navium tectum, item filtro et terra et cemento superductum. Sarraceni autem desuper iactantes copiosam stupham lini, oleum, et ignem, exusserunt instrumentum, maxime cum ita ponderosum[205] esset ut defacili retrahi non posset. Inde orta est paganis illo die maxima leticia et nobis mesticia. Intervenit etiam molestia dissensionis,[206] aliis, et maxime Flammigis,[207] volentibus dis-

[191] MS. *incundimus.*

[192] This was the second time that the camp was moved forward; see above, p. 619, lines 13–14, and p. 621, lines 17–18

[193] MS. *Portugalencium.*

[194] MS. *ossidione.*

[195] MS. *excubancium.*

[196] July 29.

[197] MS. *subsequtus.*

[198] July 30.

[199] MS. *Agglici.*

[200] This tower had evidently come into the hands of the crusaders along with the lower city.

[201] MS. *experati.*

[202] MS. apparently *is*, though the reading is not certain.

[203] MS. *Portugalencium.*

[204] August 6.

[205] MS. *poderosum.*

[206] MS. *dissesionis.*

[207] The Flemings formed no part of the original expedition on which the author sailed from the mouth of the Weser; it need therefore cause no surprise if here as elsewhere (see below, p. 632, line 3) he displays some anti-Flemish prejudice.

cedere, aliis [7v.] volentibus stare quousque caperetur civitas. Sequenti die [208] machina nostra in easdem turres ita fortiter iactibus impegit ut altera dissoluta in quadam parte corruisset. Preterea machine due regis, licet parve, satis infestabant populum intrinsecus. Proxima nocte egressus quidam Maurus detulit ad regem duo vexilla insigniora que habebant intus. Unde postea [209] luce letum festum fecerimus,[210] Mauro fiduciam prebente de capiendo opido, maxime si corrasce caperetur. In vigilia Laurencii [211] quidam miles de Galicia, qui venerat nobiscum dux [212] cuiusdam navis nostre, accessit [213] ad murum cuius pars impulsu machine corruerat et, licet imminerent et defensores turris, tamen unum angularem lapidem a muro eruit et recessit.[214] Huius [215] audacia [216] nostri provocati, ibidem turrim perforaverunt, et quod est mirabile *dictu* [217] Sarraceni adhuc desuper stabant, nec ruina domus nec sagittis celeberrimis [218] abstricti. Illa ergo vespera in studio cavandi perstiterunt; [219] sed in nocte pavore conterriti abscesserunt,[220] putantes Sarracenos, quos e vicino [221] audierunt, contra se perforantes parietem. Mane autem facto [222] lignis quibus concavitatem subpodiaverant apposito igne partem turris deiecerunt; sed sopito igne et resumpta audacia,[223] iterum cavare [224] ceperunt,[225] volentes [lapides] [226] de turri illa ex utraque parte eo modo deicere. Iterum autem appositus est ignis, et tunc tanta pars turris corruit quod nostri scalam ibidem apposuerunt et singillatim ad propugnacula ascenderunt, in quibus erat copiosa multitudo hostium; sed Domino vires nostris largiente et illis timorem immitente ita unus omnes fugavit quod rex et sui in adverso monte huius rei spectaculo

[208] August 7.

[209] Chroust emended to read *postera*.

[210] Chroust, following Gazzera, emended to read *fecimus*.

[211] August 9.

[212] Presumably a pilot, and to be compared with the *duces vie* who are mentioned above, p. 613, line 3.

[213] MS. *accesset*.

[214] MS. *rescessit*.

[215] MS. reading certain, though Chroust read *hac;* perhaps *rei* should be supplied after *huius;* cf. line 24.

[216] MS. *audatia*.

[217] This seems the necessary reading, and it is the rendering given by Gazzera and Chroust; but it cannot be made out of the manuscript, which is a puzzle.

[218] Chroust, following Gazzera, has emended to read *creberrimis*.

[219] MS. *persisterunt*.

[220] MS. *abcesserunt*.

[221] MS. *evicinos*.

[222] August 10.

[223] MS. *audatia*.

[224] MS. the passage *apposito igne . . . iterum cavere* written in margin and its proper position in the text indicated by signs.

[225] MS. *receperunt* was possibly written first and *re* erased afterwards when the marginalium was added. Chroust read *receperunt,* which is certainly incorrect as the manuscript stands.

[226] I have supplied *lapides;* Chroust emended to read *turrim illam* for *de turri illa*.

supra modum exhilarati [227] sunt et in maxima ammiratione nostram gentem laudibus extulerunt. Dei igitur Genitricis [228] virtute et non nostra, Sarraceni IIII^{or} firmissimas turres [229] et propugnacula deseruerunt, deicientes balistas et spatas perplures; sed lento gradu ad almadinam tendebant in muro per quem satis tutum habebant transsitum in utraque parte corrasce.[230] Postquam ex nostris sufficienter intraverant, paganos ad superiorem munitionem fugere [231] compulerunt. Tunc igitus murus destructus [232] est in duobus locis, spolia direpta sunt, et puteus in quo habebant fiduciam obstructus est et lapidibus et terra reppletus.[233] Illa ergo vespera nostri, licet fessi et quidam lesi, tamen communiter leti ad castra sunt reversi.

Altera die [234] cepimus fodere in terram in duobus locis ut sub[8r.]terraneum transsitum faceremus ad murum almadine [235] et eum subfoderemus; et huic operi insudatum est illo die et nocte sequenti; sed tercio die [236] egressi Sarraceni exuerunt domos sub quibus fodiebatur, et ita, succensis lignis quibus subpodiebantur antra, nostri ab opere fugati sunt; sed plures illorum a nostris sagittariis letalia exceperunt vulnera, ita ut nostri laboris dispendium eorum strage optime compensaretur. Item Flammigi ceperunt cavare murum capte urbis, qui tamen continuabatur [237] cuidam turri almadine, ita ut per concavitatem muri veniretur ad turrim; sed Sarraceni hoc considerato nocte ab opere eos fugaverunt et murum secantes a turri diviserunt. Sed quod minus [238] nocte factum fuit, postera die [239] perfecerunt, ita ut eo modo ledi non possent.

Et nota quod iam plures sed diversis temporibus a munitione ad nos confugerunt ut salvarent animas suas; et ut alios provocarent ad exeundum, nichil mali a nobis patiebantur.[240] In vigilia assumptionis beate Marie,[241] Sarracenis aggressis, nostris etiam e regione [242] aciem [243]

[227] MS. *exilarati*.
[228] MS. *Genitrice*.
[229] The four towers of *corrasce* which secured the water supply of the city; cf. above, p. 620.
[230] It would appear that a wall, with a safe passage in or along it, led up to *almadina* from each side of *corrasce*.
[231] MS. *fugare*.
[232] MS. possibly *diffractus*, as Chroust, after Gazzera, has read; but *destructus* is more probably correct.
[233] Cf. Ralph de Diceto as quoted above, p. 620, note 150.
[234] August 11.
[235] MS. *almidine*.
[236] August 12.
[237] MS. *continuebatur*.
[238] Chroust supplied *bene* after *minus*.
[239] August 13.
[240] MS. *paciebantur*.
[241] August 14.
[242] Chroust emended to read *converso*.
[243] MS. *acie*.

instruentibus, quidam Sarracenus a muro se precipitem dedit et ad nostros confugit; et cum benignissime susciperetur, sitibundus valde aquam poposcit; et siti[244] affectus fere totam faciem in ea mersit pre aviditate potandi, dicens multam hostium turbam siti interire; nam desuper in puteis non multum aque habebant, et ipsa valde salsa erat.

Proxima die post octavam Laurencii[245] totus exercitus noster armatus est et ex omni parte accessit ad murum ferens scalas, et diutino et instanti labore nisus[246] est eas erigere; sed ita crebris et gravibus iactibus est repulsus ut spes nostra cassaretur, et multi saucii reverterentur. Nonnulli etiam ex hostibus a sagittariis nostris vel perempti vel vulnerati sunt. Quidam etiam nostrorum[247] fossatum quod erat in aquilonari parte almadine[248] replere frondibus[249] moliti sunt, que[250] desuper iacto igne ilico consumpte sunt. Nec mirum si difficilis erat ascensus ad murum, cum ex una parte crepido montis esset et fossatum mirabile, ex alia parte[251] domus dense artam faciebant viam. Hac ergo iactura pavefacti Portugalenses, maxime cum cibaria ipsis et pabula[252] equis deessent, ceperunt tam regem quam nostros sollicitare de recessu. Rex etiam visus est pronus ad discedendum; sed nostri decreverunt communiter diutius hostes Christi impugnare, et hoc regi intimaverunt. Rex autem deliberato [8v.] consilio strenue cum ipsis manere consensit; et tunc rursus [resumptis animis][253] ad capiendam civitatem cepimus vigilancius insudare.

[Erecte sunt][254] autem in parte aquilonari[255] quatuor machine, una nostra, .III. regis; et ipsi[256] contra nostras $\overset{\text{or}}{\text{IIII}}$ erexerunt. Cepimus etiam fodere interea sub terram sed longius a munitione[257] ne, sicut prius, impediretur opus nostrum; quod intellexerunt pagani, et aperta porta exiverunt ut destruerent foveam; sed nostris accurentibus graviter

[244] MS. *cum;* I follow Chroust in emending to read *siti.*

[245] August 18.

[246] MS. reading certain, though Chroust, following Gazzera, incorrectly read *visus.*

[247] MS. *nostrum.*

[248] Faint traces of a dry moat high on the hill beneath the northern wall are still visible.

[249] MS. *fondibus.*

[250] MS. *qui.*

[251] MS. the passage *crepido montis . . . alia parte* written in margin; *crepido* was first written *crepito* and then corrected.

[252] MS. *popula.*

[253] MS. deteriorated from dampness at this point and the words supplied in brackets wholly illegible. Chroust, following Gazzera, printed *rursus ad capiendam* without any notice of an omission.

[254] MS. deteriorated from dampness and the words supplied in brackets almost illegible. Chroust, following Gazzera, read *Erant* for *Erecte sunt.*

[255] MS. *aquiloni.*

[256] Chroust, following Gazzera, supplied *Sarraceni* after *ipsi.*

[257] MS. *amminitione* for *a munitione.*

ex utraque parte vulnerati reversi sunt. In octava assumptionis [258] diluculo iterum exiverunt; sed non sufficienter a nostris repressi, stabant extra muros et tundebant terram si concavitas audiretur, quia timebant ad murum iam pervenisse foveam, que tamen adhuc remota erat. Quidam etiam fodiebant ut sic foveam deprehenderent; sed pauci ex nobis [259] armaverunt se ut reprimerent eos. Impetum ergo in ipsos facientes, quosdam straverunt, multis ex iaculis nostris cadentibus; et accesserunt nostri usque ad introitum porte, ita quod si omnes nostri armati essent et presto, defacili portam intrassent. Nostri ergo cum victoriosa leticia reversi sunt.

In vigilia Bartholomei [260] maxime perturbationis molestia exorta est. Rex enim Portugalensis [261] cum suis cum festinatione recedere proposuit; sed nostri vix obtinuerunt [262] ut staret adhuc $\overset{\text{or}}{\text{IIII}}$ diebus.[263] Interea ceperunt nostri fodere ab antro [264] quodam quo frumentum servabatur in terra molli et vicinius muro. In festo Barto[lomei] [265] placuit regi fovea, et suis; qui ad discedendum [266] motus erat, iterum constanti animo laborare cepit. Laboratum est ergo supra modum in fovea illa; nam cum nostri prope murum venissent, Sarraceni murum perforantes in occursum [267] nostrum [268] foveam fecerunt et ad nostros accedentes diu dimicaverunt cum eis. Tandem igneo [269] copioso fluvio, cuius materiam sollicite *com*paraverant,[270] nostros fere fugaverunt; sed tandem laboriose obstructum est foramen, et processum est in fovea nostra. Illi autem nichilominus foveam inter nostram et murum fecerunt et nostr[is (9r.) prohibuerunt ac]cessum [271] ad murum. Interius etiam longam et aliam foveam fecerunt [iuxta] [272] murum, quia credebant quod per foveam murum intrare ad eos proposuissemus; sed ipsa intentio [273] erat deicere murum. Multo autem tempore fodiendi

[258] August 22.

[259] Chroust, following Gazzera, emended to read *nostris.*

[260] August 23.

[261] MS. *Portugagalensis.*

[262] MS. possibly *obtinuere.*

[263] The consent of the king to remain was bought at the price of an alteration in the original agreement between him and the crusaders. See below, p. 631, lines 6–8.

[264] MS. reading very uncertain, possibly *aracio, aratio,* or *aratro.* I adopt the emendation of Chroust.

[265] MS. *Barto.* August 24.

[266] MS. *discendum.*

[267] MS. *ocursum.*

[268] Chroust emended to read *nostrorum.*

[269] MS. *igneus.*

[270] MS. the prefix *com* illegible.

[271] MS. deteriorated from dampness and the text supplied in brackets wholly illegible. I adopt the reading of Gazzera and Chroust.

[272] MS. deteriorated and almost illegible; Gazzera and Chroust read *prope,* but *iuxta* is more probably correct.

[273] MS. reading doubtful, possibly *incentio.*

studio detenti sumus, cotidie pugnantes in fovea cum paganis, qui similiter multiformi labore nostrum opus impedire moliebantur. Tandem die sancti Egidii [274] homines regis ad muros vocabant, de tradenda civitate tractabant. Tunc etiam plures Sarraceni fuga elapsi ad nos venerunt dicentes eos siti [275] laborare et metu foveë concuti.

Hoc modo autem convenerunt [276] pagani cum rege, ut traderent civitatem et castrum et salvi recederent cum rebus suis. Et huius rei consensum rex extorquere a peregrinis sategit, sed non perfecit. Ut autem consentirent, promisit X. milia aureorum, tandem XX. milia, que recepissemus; sed pre [277] mora que futura erat in reddendo, quia a terra sua portari necesse fuit, recusavimus. Convenimus ergo ut Sarraceni tantum cum una veste exirent, et omnia mobilia haberemus, et rex urbem. Et huic pactioni oportuit paganos obedire, quia siti defecerunt et victos cavee urgebant,[278] quia quedam turris magna quam Burge Marie [279] dicunt, id est turrim Marie, ruinosa erat propter caveam, sicut etiam vicinus murus.

Tercio quoque Nonas [280] Sebtembris [281] exivit dominus civitatis, Alb*ain*us [282] nomine, solus in equo, reliqui pedites sequebantur; sed populus noster satis turpiter quosdam exspoliavit contra pactum et

[274] September 1. Chroust, following Gazzera, supplied *Sarraceni* after *Egidii*.

[275] MS. *sui*.

[276] MS. *consueverunt*.

[277] MS. *pro*.

[278] MS. *victus cavee urgebat*, though *cavee* is hardly certain and Gazzera read it *universe;* I adopt the emendation of Chroust.

[279] Apparently the author has misunderstood, and given a fanciful interpretation to, the Moorish name of the tower which had been all but successfully undermined.

[280] MS. a mark of abbreviation erroneously placed over *Nonas* which is written in full.

[281] September 3. The date is confirmed by the *Chronicon Conimbricense*, in *Portugaliae Monumenta Historica, Scriptores*, I, p. 3 (though the edition of Flórez in *España Sagrada*, XXIII, 332, gives the year as 1190—"Era MCCXXVIII"), and also by the Anonymous of Madrid and Copenhagen (ed. A. Huici, text p. 59, translation p. 62) who says that Sancho entered Silves on Racheb 20, which corresponds to September 3, though Huici has erroneously given it as September 2. Ralph de Diceto (*Opera Historica*, II, 66) is evidently in error in giving the date as September 6. Cf. Kurth, p. 201, note 2.

[282] MS. reading uncertain; I retain the reading of Gazzera and Chroust, but *Albanius* or *Albamus* could equally well be correct. Ralph de Diceto (*Opera Historica*, II, 66) merely calls the governor of Silves the alcayde without naming him. Herculano (*Historia de Portugal*, III, 188, and note 2) has reasoned by inference from Arabic sources that this governor must have been "Abdullah, or Abu Abdullah," "son or nephew of an[earlier]*wali* of Silves"; and Kurth (p. 201, note 3), using the same sources, has tried to be even more definite. But neither Herculano nor Kurth have used the contemporary and generally well informed Arabic authority, the Anonymous of Madrid and Copenhagen (ed. A. Huici, p. 61), who names the Muslim governor at the time of the Christian conquest: "Era su gobernador entonces Aisa ben Abuhafs ben Ali, muy experimentado en la defensa de las fronteras." Presumably he was a son of Abū Ḥafṣ ʿOmar b. ʿAlī *Azannag* (of the Sanhāja tribe), one of the early disciples of Ibn Tūmart, the founder of the Almohade sect. See Ibn Abī Zarʿ, *al-Kirṭās* (*Annales Regum Mauritaniae*, ed. C. J. Tornberg, II, 153); Ibn Khaldūn, *Histoire des Berbères* (ed. Paul Casanova), II, 170. Cf. A. Bel, in *The Encyclopaedia of Islam* (ed. M. T. Houtsma and others, Leyden and London, 1913–34), *s. v. Almohades*.

verberavit.[283] Unde pene mota fuit seditio inter regem et nostros. Imminente autem nocte clausimus portam ne plures pagani exirent; et ex alia parte intraverunt nostri, et quidam primo [284] per eandem portam, et fuerunt cum paganis tota nocte, et pagani claudebantur in domibus. Quidam etiam contra pactum torquebantur pro pecunia monstranda.[285] Mane [286] modestius educti sunt de tribus portis; et tunc primo vidimus defectum eorum, nam macilentissimi erant et vix gradiebantur. Multi reptabant, alii per nostros sustentabantur, alii in plateis iacebant vel mortui vel semivi*vi*; [287] et fetor maximus erat tam de cadaveribus hominum quam ani[9v.]malium brutorum in civitate.

Captivi autem Christiani efferebantur vix spirantes; nam, sicut nobis retulerunt, intra IIII (or) dies unus non habebat nisi tantum aque quantum testa ovi capere poterat, et aliqui minus; et nulli prorsus dabatur aqua nisi pugnare volenti, et ipsi modicum quid; dividebant autem cum [288] uxore et filiis. Nec panis fiebat propter defectum [289] aque vivificus; [290] sed comedebant ficus, et ideo reservata est maxima copia annone. Captivi etiam denudabantur noctibus et iacebant super frigidos lapides ut sic humectarentur et viverent. Comedebant etiam mulieres et pueri humidam terram. Et notandum quod primo, cum venimus, Silvia habebat quadringentos et quinquaginta captivos, sed vix invenimus vivos ducentos.

De habitatoribus autem cum traderetur civitas erant promiscui sexus .XV. milia et octingenti. Nota a die quo obsedimus usque ad diem quo capta est civitas fluxerunt VI. ebdomade et tres dies.[291] Preterea Silvia multo [292] municior erat quam Ulixbona et in decuplo locuplecior et edificiis preciosior. Asseruerunt etiam Portugalenses quod in Ispaniis munitior non esset civitas, et Christianis tam infesta.

Sciendum etiam quod toto tempore obsidionis Portugalenses nec laborabant nec pugnabant, sed tantum insultabant nobis quod in

[283] MS. *verberisavit.*

[284] MS. reading certain, though without any warrant Gazzera read *vero* and Chroust, *Portugalenses*, thereby completely altering the meaning. It should be noted that on p. 631, lines 5–6, below, the author says specifically: "Capta civitate soli nos Franci possedimus eam, et nulli alii concedebatur introitus."

[285] MS. *mostranda.*

[286] September 4.

[287] MS. deteriorated from dampness and the end of *semivivi* illegible.

[288] MS. *cum* inserted above the line.

[289] MS. *deffectum.*

[290] MS. *tuificus.*

[291] From July 21, on which the siege operations were actually started, through September 3, on which the exodus of the Saracens from the city began, was exactly six weeks and three days.

[292] MS. *multa.*

vanum laboremus et quod inexpugnabilis esset munitio.[293] Ipsum etiam regem inducebant ut discederet et nobis discessum multiformiter persuaderet. Maxima etiam pars nostrorum desperans abire volebat, sed Deus misericorditer et [294] mirabiliter nos tam felici consummationi conservabat. Nota noster exercitus [295] tantum habebat, cum [296] primo venimus, tria milia et quingentos cuiuslibet ordinis vel etatis viros, vel paulo pauciores.[297] Exercitus [298] autem regis multus erat equitum, peditum et galiotorum;[299] et erant cum eo milites religiosi de tribus sectis: Templarii, Iherosolimitani milites qui ferunt gladios in vestibus, qui ducunt [300] uxo[10r.]res, et assidue movent guerram cum Sarracenis, et tamen regulariter vivunt.[301] Item milites de ordine Cisterciensi, qui tantum eam indulgenciam habent quod carnibus vescuntur III. diebus in epdomada, sed una vice et uno ferculo [302] cum domi sunt, sed cum in expeditione sicut reliqui homines;[303] quorum caput est Callatravia [304]

[293] Such an attitude of disparagement towards the Portuguese is a commonplace among northern writers of this epoch.

[294] MS. *ita;* I adopt the emendation of Chroust.

[295] MS. *excercitus.*

[296] MS. *tum.*

[297] The number is concurred in by Ralph de Diceto (*Opera Historica,* II, 66). In view of the number of ships in which the northern forces arrived at Silves (see above, p. 618, note 125, there is no reason to doubt the substantial correctness of this figure. A careful estimate of King Richard's forces on the Third Crusade as they sailed for the Mediterranean from the mouth of the Tagus in 1190 indicates that his ships (some of which may have been unusually large) carried on an average about eighty crusaders and a crew of about twenty-five. See B. N. Siedschlag, *English Participation in the Crusades, 1150–1220,* pp. 145–148.

[298] MS. *excercitus.*

[299] MS. *galietarum.*

[300] MS. *ducant.*

[301] *Die ursprüngliche Templerregel,* ed. Gustav Schnürer, in *Studien und Darstellungen aus dem Gebiete der Geschichte* III, Heft 1 (Freiburg im Breisgau, 1903), p. 148: "DE CONIUGATIS. Fratres autem coniugatos hoc modo habere vobis permittimus, ut, si fraternitatis vestre beneficium et participationem unanimiter petunt, uterque substantie sue portionem et quicquid amplius acquisierint, unitati communis capituli post mortem concedant et interim honestam vitam exerceant et bonum agere fratribus studeant; sed veste candida et chlamide alba non incedant. Si vero maritus ante obierit, partem suam fratribus relinquat et coniux de altera vite sustentamentum habeat. Hoc enim iniustum consideramus, ut cum fratribus Deo castitatem promittentibus fratres huiusmodi in una eademque domo maneant." According to Schnürer this rule dates from 1130.

[302] MS. *ferculo* written in margin and very faintly legible.

[303] In a bull of Alexander III (September 25, 1164), addressed to García, the master, and to the brothers of Calatrava, and approving the foundation of their order, the pope says (Migne, *Patrologia Latina,* CC, col. 311): "Praeterea, ea quae de victu et vestitu vestro praedicti abbas et fratres Cistercienses, et universum capitulum ejusdem ordinis a vobis tenenda sanxerunt, vobis auctoritate apostolica confirmamus . . . Tribus vero in hebdomada diebus, id est feria tertia, quinta, et Dominica, cum praecipuis diebus festis, carnibus vesci licebit; uno tantum ferculo, et unius generis, quantum ad carnes pertinet, contenti eritis." In the confirmations of Gregory VIII and Innocent III, in 1187 and 1199 (*Diffiniciones de la Orden y Cavalleria de Calatrava conforme al Capitulo General celebrado en Madrid Año de M.DC.LII,* Madrid, 1661, pp. 39–51; Migne, *Patrologia Latina,* CCXIV, col. 590–593), there is added: "Qui autem in castris militiae fuerunt, pro magistri arbitrio jejunia observabunt."

[304] It is not surprising that the author should identify the military order of Calatrava with the Cistercians, in view of the close connection between them which is indicated by the papal

in regno Castelle et Ebora in regno Portugalensi, sed Callatravia mater est et Ebora filia.[305] Item Iherosolimitarum alii erant de Templo, alii de sancto Sepulcro,[306] alii de Hospitali; et singuli habent reditus in terra illa.

Capta civitate soli nos Franci[307] possedimus[308] eam, et nulli alii concedebatur introitus. Ex prima enim convencione nostra erant omnia mobilia, sed cum assidue sollicitarent nos [de][309] discessu[310] Portugalenses, dedimus eis partem, sed arbitrio regis taxandam.[311] Possessa ergo munitione rex nitebatur a nobis impetrare annonam, que copiosa[312]

letters cited in note 303 above, all of which are addressed "fratribus ordinis de Calatrava . . . secundum ordinem Cisterciensium fratrum viventibus." Pope Alexander III, writing [1179–81] to the king of Portugal, says, ". . . cum fratres de Calatrava sint ordine Cisterciensium professi . . . ": Carl Erdmann, *Papsturkunden in Portugal*, No. 81, in Gesellschaft der Wissenschaften zu Göttingen, *Abhandlungen*, Phil.-Hist. Class, new series, XX (1927), part 3, p. 254. On the whole subject see Hans Prutz, *Die Geistlichen Ritterorden* (Berlin, 1908), pp. 74–79.

305 The military order of Évora is better known by its later name, the Order of Avis. Its identification with the Cistercians in the present text is again not surprising in view of the close connection between them which is fully illustrated by the early statutes of Avis which are printed in Portuguese in *Regra da Cavallaria e Ordem Militar de S. Bento de Avis* (Lisbon, 1631), fol. 2–3. The relation of Évora, or Avis, to Calatrava is shown by a bull of Innocent III, dated May 17, 1201, which is printed in Portuguese translation, *ibid.*, fol. 3, and is addressed "Aos muito amados filhos Mestres, & Irmãos da Milicia d'Euora, q̃ professais a Ordẽ de Calatraua." For the best modern account of the early history of the order of Évora or Avis and its relation to the Cistercians and to Calatrava, see H. Prutz, *Die Geistlichen Ritterorden*, pp. 89–91; also, especially for the citation of important documents, Silva Lopes, *Relação*, pp. 83–88.

306 Notwithstanding the existence of a large, though unsatisfactory, literature on the subject (see F. Pasini-Frassoni, C. A. Bertini, and C. de Odriozola, *Histoire de l'Ordre Militaire du Saint-Sépulcre de Jérusalem*, Rome, [1908], with extensive bibliography), the present text offers perhaps the only significant (though, to be sure, not quite specific) early evidence of the existence of a military order of the Holy Sepulchre, and as such it must probably be regarded not as a demonstration but rather as due to some misunderstanding on the part of the author. Such a misunderstanding is hardly surprising in view of the fact that the canons of the Holy Sepulchre had on more than one occasion been designated along with the Templars and the Hospitallers as recipients of donations in the Iberian Peninsula. For example, King Alfonso I of Aragon and Navarre, apparently having failed in an earlier plan to establish a specifically Spanish military order (Marquis d'Albon, *Cartulaire Général de l'Ordre du Temple*, Paris, 1913, No. VI; J. von Aschbach, *Geschichte Spaniens und Portugals zur Zeit der Herrschaft der Almoraviden und Almohaden*, Frankfort, 1833–37, II, 14), bequeathed in a famous will (October 1131) his whole realm in three equal parts to "Sepulcrum Domini quod est Iherosolimis et eos qui observant et custodiunt illud et ibidem serviunt Deo, et Ospitale pauperum quod Iherosolimis est, et Templum Domini cum militibus qui ad defendendum christianitatis nomen ibi vigilant" (Albon, *op. cit.*, No. XL). The terms of this will were never carried out, but in the settlement which was arranged in 1140 between the legatees and Alfonso's successor (*Coleccion de Documentos Inéditos del Archivo General de la Corona de Aragon*, ed. P. de Bofarull y Mascaró and others, Barcelona, 1847–1910, IV, No. XXXII) and which was confirmed by the papacy in 1158 (*ibid.*, No. CXXX), the interests of the Sepulchre were treated upon equal terms with those of the Hospital and the Temple. Cf. J. Delaville Le Roulx, *Les Hospitaliers en Terre Sainte et à Chypre, 1100–1310* (Paris, 1904), pp. 47–49. Attention may also be called to a private donation of October, 1131, which is printed in Albon, *op. cit.*, No. XXXIX.

307 MS. reading somewhat doubtful because of unusual abbreviation.

308 MS. *possidemus*.

309 I follow Chroust in supplying *de* which the sense seems to require.

310 MS. *dicessu*.

311 Cf. above, p. 617, note 102.

312 MS. a superfluous mark of abbreviation above *copiosa* which is written in full.

erat, et omnibus aliis multo melior, pro porcione sua.[313] Sed cum prohibuissemus ne aliquid ferretur de civitate, ut in ipsa divideremus predam, nostri quidam, et maxime Flammigi, furtive vendebant ultra muros frumenta Portugalensibus. Unde rex valde commovebatur, asserebat namque melius esse non fuisse captam urbem quam ammittere pro penuria panis; et in ipsa commotione nostri sine consensu magistratuum et contra pactum [314] efferebant predam ante distributionem inter Portugalenses et nos faciendam. Unde nos, ne mine regis in dampnosas lites convalescerent, reddimus ei urbem adhuc [315] opibus plenam, poscentes ut sicut regiam maiestatem deceret, nobiscum impertiretur, considerato tam labore nostro quam dampno. Rex vero omnia sibi vendicans nichil nobis reddidit; [316] et ideo peregrini sic iniuriose tractati minus amice ab eo separati sunt. Preterea, antequam caperetur urbs, decimam partem tocius terre voverat sepulcro Domini pro nostra exhortatione, ut [317] mora nostri obsequii hoc munere compensaretur; sed post captam urbem votum non implevit.[318]

In vigilia nativitatis Virginis Marie [319] naves conscendimus et lente versus mare processimus. Rex vero V$\overset{o}{\mathrm{I}}$ die [320] [10v.] *abinde*,[321] rebus dispositis, et urbem principe milicie sue [322] et multis militibus *muniens*,[323] ad propria rediit. Nos autem moram in portu fecimus [324] [tum] [325]

313 MS. *suarum;* I adopt the emendation of Chroust.

314 MS. *compactum;* I follow Gazzera and Chroust in emending to read *contra pactum.*

315 MS. *ad hoc;* I adopt the emendation of Gazzera and Chroust.

316 The passage cannot be taken to mean that the crusaders received no part of the spoils of the city, since in the next paragraph they are shown dividing spoils among themselves before leaving port. Cf. Silva Lopes, *Relação*, pp. 90–92.

317 MS. *ubi.*

318 Cf. Silva Lopes, *Relação*, pp. 90–93.

319 September 7.

320 September 12 (?). Kurth (p. 205), perhaps by a slip, dated the king's departure September 11; Chroust, apparently interpreting *sexto die* as equivalent to *feria sexta*, dated it September 8.

321 MS. reading uncertain because of deterioration. Gazzera and Chroust read *abunde.* It should be noted that the present condition of the manuscript renders the reading of the next twenty lines of the text uncertain at many points, and much reliance has had to be placed upon the readings of Gazzera in whose time the manuscript was presumably in a somewhat more legible condition.

322 It is no more possible to identify the Portuguese *princeps militiae* here than on p. 618 above. Kurth (p. 184, note 2, and p. 204) has argued that the commander here referred to is different from the one mentioned in the earlier passage and has asserted that he was Alvaro Martins; but the charter of King Sancho I on which he relies (Bernard de Brito, *Monarchia Lusytana*, [Alcobaça] and Lisbon, [1597]–1727, IV, 259–260) proves no more than that before July 27, 1190 (the date of the charter) Alvaro Martins, the king's vassal, had been killed by the Saracens at Silves.

323 MS. reading doubtful because of deterioration.

324 They did not leave port until September 20. See below, p. 635, line 8.

325 MS. here and four words farther on nearly illegible; the reading *tum . . . tum* is Gazzera's.

pro divisione spoliorum [tum] pro reparatione duarum navium fractarum. Interea princeps milicie regis assumpsit [quendam] [326] clericum Flammigum [327] ad episcopatum Silvie, et cum ipso mansserunt aliquot Flammigi. Sollicitavit etiam per eum peregrinos ut [cum ipso irent] [328] ad opidum [obsidendum] [329] ad unam dietam distans quod tam pagani quam [Christiani Sanctam Mariam] [330] de Pharrum [331] vocant, sed communem assensum extorquere non potuit.

Hec autem sunt [castel]la [332] que sortita est Christianitas per aquisitionem Silvie: [Carph]anabal,[333] [Lagus],[334] Alvor,[335] [Por]ci-

[326] MS. almost illegible; the reading is Gazzera's.

[327] Cf. Ralph de Diceto, *Opera Historica*, ed. Stubbs, II, 66: "Civitate tandem sordibus et ydolatriis emundata, die nativitatis beatae Mariae [8 Sept.] Portugalensis episcopus Macomeriam in honore beatae Virginis dedicavit, praeficiens ibidem episcopum qui de partibus Flandriae causa peregrinationis advenerat." The bishop's name was Nicolaus, as a number of documents show. In December 1189 King Sancho I made a donation of the castle of Alvor to the monastery of Santa Cruz of Coimbra which "Niculaus Syluensis episcopus confirmat" (Brito, *Monarchia Luystana*, IV, fol. 15); in February, 1191, the king made a donation to the monastery of Alcobaça, and among the bishops attesting the charter appears "Niculaus Syluensis" (*ibid*, fol. 259–260); in March, 1190, Bishop Nicolaus himself, at the king's request, made a donation of the *ius ecclesiasticum* of the castle of Lagos to the monastery of Saint Vincent de Fora at Lisbon, the king himself having already given the castle (Nicolao de Santa Maria, *Chronica da Ordem dos Cónegos Regrantes do Patriarcha S. Agostinho*, Lisbon, 1668, II, 129), and among the canons of Silves who consent to this donation are "William, the dean, Peter, the treasurer, and Lambert, the archdeacon," some of whom, at least, may well have been fellow Flemings who remained at Silves with Nicolaus when the rest of the crusaders departed. See Fortunato de Almeida, *História da Igreja em Portugal* (Coimbra, 1910–24), I, 189–190; J. B. da Silva Lopes, *Memorias para a Historia Ecclesiastica do Bispado do Algarve* (Lisbon, 1848), pp. 126–137; *idem, Relação*, pp. 93–99.

[328] MS. *cum ipso irent* now illegible; the reading is Gazzera's.

[329] I supply *obsidendum* which the sense seems to require.

[330] MS. *Christiani Sanctam Mariam* illegible; the reading is Gazzera's.

[331] Faro, some thirty miles southeastward of Silves. It is the principal city and the administrative centre of Algarve, and has been an episcopal see since the sixteenth century when it succeeded Silves. The promontory on which it stands, which is the southernmost extremity of Portugal, is still known as Cabo de Santa Maria. For a miraculous explanation of the Muslim name Santa Maria (*Shantamariyat*) see *Gesta Regis Henrici Secundi* (ed. William Stubbs), II, 121. For the origin of the name Faro, from the *Beni Ḫarun* who ruled there in the eleventh century, see David Lopes, "Toponymia Arabe de Portugal," in *Revue Hispanique*, IX (1902), 63–74.

[332] MS. *castella* now partly illegible; the reading is Gazzera's.

[333] MS. *Carphanabal* now partly illegible; the reading is Gazzera's. The modern identification is very uncertain, though it is possible that Cape St. Vincent is meant. An extract from an apparently as yet unpublished charter of Prince Henry the Navigator reads as follows: ". . . no outro cabo que ante o dito cabo de Sagres está aos que veem do ponente para levante que se chamava Terça-Nabel, ao qual puz nome Villa do Infante." (There appear to be two copies, the Marquez de Souza Holstein, in *Conferencias celebradas na Academia Real das Sciencias de Lisboa*, First Conference, 1877, p. 24, taking it from a copy in the Biblioteca Nacional of Lisbon, and Silva Lopes, *Relação*, p. 100, deriving it at second hand from the Archivo Nacional da Torre do Tombo.) There is other evidence (see Jaime Cortesão, in *Guia de Portugal*, II, 318–319) that Prince Henry's "Villa" was located on Cape St. Vincent, rather than at Sagres as has commonly been supposed, and the quotation seems to show that before the prince established his "Villa" the site was called *Terça-Nabal*, which has enough likeness to Carphanabal to suggest a possible identification. Silva Lopes (*op. cit.*, pp. 99–102), with the quotation before him, unaccountably made an identification with Sagres and in so

munt,[336] Munchite,[337] Montagut,[338] Cab[oiere,[339] Sanctus Bartholomeus de Mussiene],[340] Paderne.[341] Hec omnia subiacebant dominio Silvie, [que] prorsus va[cua r]eli[quimus] [342] sed firma satis et edificata, quorum habitatores quidam in Alvor [343] a precedentibus occisi sunt, sed maxima pars confugerat in Silviam. Et Albafere [344] se tradidit regi

doing has been followed by Chroust. I know of no early evidence for a name at all resembling *Carphanabal.*

[334] MS. *Lagus* now illegible; the reading is Gazzera's. The port of Lagos, situated some sixteen miles east of Cape St. Vincent, supposed to be *Lacobriga* of antiquity and *al-Zāwiya* or *Halk al-Zāwiya* of the Arab geographers. Cf. *Guia de Portugal,* II, 299–301; Idrīsī, p. 218.

[335] Alvor, at the mouth of the small river of the same name, some five miles east of Lagos, supposed to be *Portus Hannibalis* of antiquity, and apparently called "The Port" by the Anonymous of Madrid and Copenhagen. Cf. *Guia de Portugal,* II, 280; F. X. d'A. Oliveira, *A Monografia de Alvor* (Oporto, 1907), pp. 37–43; *El Anónimo de Madrid y Copenhague,* ed. Huici, p. 61.

[336] MS. *Porcimunt* now partly illegible; the reading is Gazzera's. Portimão, forerunner of the modern port and industrial town of Villa Nova de Portimão, some three miles east of Alvor at the mouth of the Arade River which flows down from above Silves. Cf. Silva Lopes, *Relação,* pp. 104–105; *Guia de Portugal,* II, 270–271; *Enciclopedia Universal Ilustrada, s. v. Villa Nova de Portimão.*

[337] "Nunc vicus Monchigud," according to Gazzera, who has been followed by Kurth (p. 205) and by Chroust, but apparently no such place exists in the region, and it seems quite certain that the true identification (as proposed by Silva Lopes, *Relação,* p. 104) is with Monchique in the Serra de Monchique some fifteen miles north of Villa Nova de Portimão. There was a castle there in the middle of the twelfth century which was known to the Muslims as Morjīk. See C. F. Seybold, "Monchique et Arrifana d'Algarve," in *O Archeologo Português,* VIII (1903), 123–124, and appendix A, p. 653, note 51, below. The well-known modern watering place of Caldas de Monchique lies a little to the southward. Cf. *Guia de Portugal,* II, 284–298.

[338] "Hodie Montagudo pagus in monte Jorge-Alboniz," according to Gazzera, who has been followed by Kurth (p. 205) and Chroust. No such modern place appears to exist in the region, but Silva Lopes (*Relação,* p. 105) explains: "Na frequezia de Santo Estevão ha hum sitio chamado Mont-agudo, que tem ao presente 38 fogos espalhados por casaes, e consta por tradição ter existido hum casal com o mesmo nome no cimo do serro de Jorge Moniz, do qual ainda apparecem alguns vestigios. Fica este morro ao N. da igreja não menos d'hum quarto de legoa declinando para N. O. proximo á ribeira da Aceca [Asseca]." However, the parish of Santo Estevão is only three or four miles west of Tavira and nearly fifty miles east of Silves, and it may well be doubted whether it should be included in a list of places which fell into Christian hands as a result of the conquest of Silves. The identification must therefore remain uncertain.

[339] MS. *Caboiere* now partly illegible; the reading is Gazzera's. Probably to be identified with Carvoeiro some six miles to the south of Silves or with a fortress on near-by Cabo de Carvoeiro. This identification, first made by Gazzera, has been accepted by Kurth (p. 105) and by Chroust, but Silva Lopes (*Relação,* pp. 105–107) questioned it, since he found another site called Carvoeiro (*sitio chamado Carvoeiro*) in the parish of São Bartolomeu de Messines. I have failed to identify any such site.

[340] MS. now wholly illegible. Gazzera read *Mussiene* only, but from the space available in the manuscript it seems not improbable that the reading should be *Sanctus Bartholomeus de Mussiene.* São Bartolomeu de Messines, situated some ten or eleven miles northeast of Silves. Cf. *Guia de Portugal,* II, 218–220.

[341] Paderne, some eight miles southeastward from São Bartolomeu de Messines and almost due east of Silves. Cf. *Guia de Portugal,* II, 220.

[342] MS. *que prorsus vacua reliquimus* now partly illegible, as indicated by brackets in the text; the reading is Gazzera's.

[343] Gazzera's reading, *quia habitatores oppidi Alvor,* which has been accepted by Chroust, is certainly erroneous.

[344] Albufeira, on the south coast of Portugal, nearly midway between Villa Nová de Portimão and Faro. Cf. *Guia de Portugal,* II, 221–225.

pre nostro timore, cuius opes in Silviam transtulit. Et nota a Silvia usque Ulixibonam septem diete sunt,[345] inter quas tuta non fuit habitatio nec Christianis nec Sarracenis pro utriusque gentis [di]scursu;[346] sed nunc tutissimam Christiani habent manssionem in felicissima terra dum possident Silviam. Et nota quod post VIII° dies a redditione urbis[347] cecidit maxima pars muri quem prius nostri subfoderant, et ipsi Sarraceni nostris[348] fossoribus occurentes.

In vigilia Mathei[349] portum Silvie exivimus, a sinistris relinquentes Sanctam Mariam de Pharrum et Thaviram.[350] In die Mauricii[351] mane venimus contra Allem[ir][352] que preterfluit Siviliam.[353] Distat autem Sivilia[354] a mari duabus dietis, civitas opulentissima et maxima. Item a Sivilia ad tres dietas sita est Corduba super eundem fluvium. A Silvia[355] autem ad Odianum,[356] quendam fluvium, sunt diete [ ][357] terre culte habentis hec opida: Pharum, Lole,[358] Castala',[359] Taviram,

[345] Idrīsī (pp. 219, 222) gives the distance as a journey of six days—four from Silves to Alcácer do Sal, and two from there to Lisbon.

[346] MS. *discursu* now partly illegible; the reading is Gazzera's.

[347] September 10.

[348] MS. *nostri.*

[349] September 20.

[350] Tavira, ancient *Balsa*, on the south coast of Portugal some nineteen miles eastward from Faro and about fourteen miles from the Spanish border. Cf. *Guia de Portugal*, II, 255–258.

[351] September 22.

[352] MS. reading of the river name uncertain at the end, but the clearness of the reading where the word recurs below, p. 637, makes it very probable that *Allemir* is the correct form; Chroust read *Allemur*. The Guadalquivir, ancient *Baetis*, called by Idrīsī the great river, or the river of Cordova.

[353] MS. *Silviam.*

[354] MS. *Silvia.*

[355] Gazzera read *Sivilia* without warrant in the manuscript; Chroust gave the correct reading in a note but erroneously emended the text to read *Sivilia*. However, there is no doubt that *Silvia* is both the correct and the necessary reading, since all the places mentioned in the ensuing list lie to the west of the Guadiana, except Serpa which is only five or six miles to the east of that river and still far within the borders of modern Portugal, and Loulé and Faro are not more than twenty-five or thirty miles eastward from Silves. Of the region between the Guadiana and Seville the author goes on to speak in the next sentence.

[356] The Guadiana River.

[357] There is evidently an omission in the manuscript at this point. It is possible, as Chroust has suggested, that a whole line has been dropped, but it is perhaps more probable that nothing more than a numeral is wanting. There is no warrant in the manuscript for Gazzera's reading, *tres* for *terre culte* which follows the omission.

[358] Loulé, about ten miles northwest of Faro. Cf. *Guia de Portugal*, II, 226–229.

[359] MS. reading of this place name uncertain at the end, and I have not extended the abbreviation. Gazzera and Chroust read *Castalar* and identified it as "Locus quondam situs in praealta rupe ad flumen Alcaria, cuius rudera adhuc appellantur Castellos." Though Silva Lopes (*Relação*, pp. 107–108) remained in doubt, it is certain that the correct identification is with Cacela, as Herculano (*Historia de Portugal*, 8th ed., III, 344–345) perceived with the aid of Idrīsī who called it *Castella*. Idrīsī (pp. 216–217) describes it as "une forteresse construite sur les bords de la mer" and locates it between the mouth of the Guadiana and Tavira, fourteen miles from the latter. Actually it is situated on a hill hard by the coast and six or seven miles northeastward of Tavira, about midway between Tavira and the Guadiana. Cf. *Guia de Portugal*, II, 258; U. S. Hydrographic Office, *East Atlantic Pilot*, p. 182. On the transformation

Mertulam,[360] Serpam;[361] que facile cepissemus, et terram illam, que Algarbia[362] dicitur, integraliter Christianis possidendam reliquissemus, si odium regis et quorundam nostrum[363] execranda festinatio non prohibuisset. Ab Odiana usque Siviliam terra est prorsus sterilis et deserta, duas habens dietas. Tantum unum opidum est in ripa maris, nomine Saltes,[364] quod pre metu peregrinorum incole deseruerant; et confugerunt ad montana, ad castrum nomine Elva;[365] quod est in strata vie terrestris de Odiana ad Siviliam tendentis,[366] in qua etiam sunt Nebula[367] et Fealcazar,[368] castra fortia. De Sivilia in stra[11r.]ta versus strictum

of *Castella* into *Cacela* see David Lopes, "Toponymia Arabe de Portugal," in *Revue Hispanique*, IX (1902), 39–40.

[360] Mértola, ancient *Myrtilis*, on the Guadiana at the upper limit of navigation, some thirty-six miles from its mouth. It was a stronghold of great importance in the disordered period of the decline of the Almoravides (see Francisco Codera, *Decadencia y Desaparición de los Almoravides en España*, Saragossa, 1899, pp. 33–53 *et passim*), and it is not surprising that Idrīsī (p. 217) describes it as "si connue par la bonté de ses fortifications." It is also noticed by the author of the *Gesta Regis Ricardi Secundi* (ed. Stubbs, II, 122); and by Roger of Howden (*Chronica*, ed. William Stubbs, London, 1868–71, III, 47). Cf. *Guia de Portugal*, II, 162–164.

[361] Serpa, some twenty or more miles north of Mértola and several miles to the east of the Guadiana. Cf. *Guia de Portugal*, II, 166–168.

[362] Algarve, from Arabic *al-Gharb*, meaning the West. Once the name of the whole southwestern part of the Iberian peninsula, it still survives as that of the southernmost province or district of Portugal, with its administrative centre at Faro. Cf. C. F. Seybold in *Encyclopaedia of Islam*, *s. v. Algarve*.

[363] Chroust, following Gazzera, has emended to read *nostrorum*.

[364] MS. reading clear; there is no warrant for *Saltis*, the reading of Chroust. The mediaeval community has vanished and only the island remains. "Saltes Isle, opposite the entrance to the Tinto River, is nearly 2.5 miles in length . . . and 1 mile in width. It is very low, sandy, and is partly covered with low brushwood and partly cultivated": U. S. Hydrographic Office, *East Atlantic Pilot*, p. 188. Roger of Howden (*Chronica*, ed. Stubbs, III, 47) mentions a castle there; but the fact that the inhabitants fled at the approach of the crusaders as they formerly had done at the approach of other northern raiders, makes it seem likely that it was still largely unfortified, as it was in the time of Idrīsī, whose description (p. 216) may well be quoted: "Quant à la ville de Chaltîch, elle n'est point entourée de murailles, ni même d'une clôture. Toutefois les maisons y sont contiguës; il y a un marché. On y travaille le fer, sorte d'industrie à laquelle on répugne ailleurs de se livrer parceque le fer est d'un travail difficile, mais qui est très-commune dans les ports de mer, dans les lieux ou mouillent les grands et lourds bâtiments de transport. Les Madjous [Normans] se sont emparés à plusieurs reprises de cette île; et les habitants, chaque fois qu'ils entendaient dire que les Madjous revenaient, s'empressaient de prendre la fuite et de quitter l'île."

[365] Huelva, ancient *Onuba*, the present capital of the Spanish province of Huelva. It is situated on the left bank of the Odiel River above its junction with the Tinto and several miles north of the island of Saltes. The modern town is located on the southern slope of hills which rise to a height of 120 to 140 feet. The author's use of the word *montana* is therefore justified. Idrīsī (p. 216) also refers to "une montagne au-dessus de laquelle est la ville d'Huelba" and describes the city as "ceinte d'une muraille en pierres, pourvue de bazars où l'on fait le négoce et ou l'on exerce divers métiers."

[366] The modern highway, like the ancient, runs directly from the mouth of the Guadiana through Huelva to Seville.

[367] Niebla, ancient *Ilipla*, on the right bank of the Rio Tinto, some twenty miles above Huelva at the point where both the highroad and the railroad to Seville cross the river. See the interesting description of Niebla by Idrīsī (p. 215). Cf. E. Lévi-Provençal in *Encyclopaedia of Islam*, *s. v. Niebla*.

[368] Possibly, but by no means certainly, to be identified with Aznalcázar some fifteen miles southwestward from Seville. It is on a road as well as on the railroad from Seville to Niebla

mare [369] sunt opida Scheres,[370] Roda,[371] Cadiz, [Algazin]ir.[372] De Allemir [373] usque Iezitarisif,[374] quod est opidum iuxta caput [375] stricti maris, est dieta et dimidia. A dextris ultra mare reliquimus Affricam, terram planam et optimam, usque ad strictum mare; et Phadala [376] prima civitas [oc]currit [377] que est e regione Sancte Marie de Chaphairum;[378] [item La]bu,[379] Anaphe,[380] Zale,[381] *In*zemitz,[382] Methena,[383]

and Huelva, but it is not on the modern main highway. This is presumably the identification which Gazzera intended to give, though he spelled the name Azialcacar. Chroust proposed as an alternative Aznalcollar some fifteen or twenty miles to the northward.

[369] The Strait of Gibraltar.

[370] Jerez de la Frontera on the main road from Seville to Cádiz and the Strait and some ten miles from the sea. Cf. *Enciclopedia Universal Ilustrada, s. v. Jerez de la Frontera.*

[371] Rota, on the coast at the north side of the entrance to Cádiz Bay. See U. S. Hydrographic Office, *East Atlantic Pilot*, pp. 207–208. A road leads directly from Jerez to Rota, but the present land route to Cádiz runs around the bay to the eastward.

[372] MS. *Algazinir* now largely illegible; the reading is Gazzera's; Chroust read *Algaziuir* or *Algazimir*. Apparently Algeciras, opposite Gibraltar on the west side of Gibraltar Bay; cf. below, p. 640, line 12, where the author gives the name much more correctly; also *ibid.*, note 423; C. F. Seybold, in *Enciclopedia of Islam, s. v. Algeziras.*

[373] MS. reading very nearly certain, though Chroust read *Alleuur* or *Allemin* and emended the text to read *Allemur*. The Guadalquivir River; see above, note 352.

[374] Tarifa, from Arabic *Jazīrat Tarīf*, "island of Tarif," the name being derived from Tarif, the leader of the first Muslim forces to land in Spain in 710 A.D. "Tarifa, formerly an islet, is now a small peninsula, joined to the mainland by an artificial causeway," says the *Mediterranean Pilot*, I (3rd ed., Washington, 1930), 67. It is at the southernmost extremity of Spain and some fifteen miles southwest of Algeciras. Cf. E. Lévi-Provençal, in *Encyclopaedia of Islam, s. v. Tarif* and *Tarifa*. I can find no justification in modern usage for *Jeira-Tarifa*, the form of the name given by Chroust.

[375] MS. *capud.*

[376] Fedala, French Morocco, on Cape Fedala, thirty-four miles southwestward of Rabat and fifteen miles northeastward of Casablanca. Its position is well described as opposite (*e regione*) Cabo de Santa Maria de Faro on the south coast of Portugal. Cf. U. S. Hydrographic Office, *East Atlantic Pilot*, p. 382. Idrīsī (p. 83) comments on its active commerce with Spain.

[377] MS. *occurrit* partly illegible due to a smear of ink of modern origin.

[378] Evidently Faro on the south coast of Portugal which appears as Santa Maria de Pharrum on p. 633, line 6, and p. 635, line 9, above.

[379] MS. *item Labu* all but obliterated by a smear of ink of modern origin; the reading is Gazzera's. It has proved impossible to identify this place. Gazzera referred to a "flumen nomine Lebu" by which he perhaps meant the Wadi Sebu, a considerable river which rises in the Atlas Mountains and empties into the Atlantic at Mehediya (al-Mahdīya) some seventeen miles northward of Rabat (see U. S. Hydrographic Office, *East Atlantic Pilot*, p. 379), but the reference is to a place, not to a river.

[380] Dar el-Beida or Casablanca, the *Anfa* of Idrīsī (p. 84), on the Atlantic coast of French Morocco some two hundred miles southwest of Tangier. Cf. U. S. Hydrographic Office, *East Atlantic Pilot*, pp. 383–386; G. Yver, in *Encyclopaedia of Islam, s. v. Dār al-Bēḍā.*

[381] Salé, on the Atlantic coast of French Morocco, opposite Rabat on the right bank of the Wadi Bu Ragrag. Cf. U. S. Hydrographic Office, *East Atlantic Pilot*, p. 380; H. Basset, in *Encyclopaedia of Islam, s. v. Sala.* On its situation and flourishing commerce see Idrīsī, p. 85.

[382] MS. reading somewhat uncertain; Chroust read *Tuzemitz*, which is possible, but *Inzemitz*, the reading of Gazzera, seems more probably correct. Evidently Azemmur, a town on the coast of French Morocco some thirty-five miles southwest of Casablanca, on the left bank of the Wadi Umm al-Rabia near its mouth. Cf. U. S. Hydrographic Office, *East Atlantic Pilot*, p. 386; G. Yver, in *Encyclopaedia of Islam, s. v. Azemmūr.*

[383] Identification uncertain; possibly the "great town of Masina . . . formerly surrounded with walls and provided with markets, but . . . actually in ruins," which Idrīsī (p. 203) locates on the Wadi Sebu. Mehediya, the proposed identification of Chroust, seems to have very little in its favor.

Azila,[384] Tanchia,[385] que est in capite stricti maris. Marrocus [386] autem metropolis Affrice [387] est in eadem plana terra, sed distat a mari quinque dietis. A capite stricti maris in ulteriori parte incipiunt montana valde alta; et dicitur terra illa montuosa Agummera [388] vel Barbaria; [389] et durat usque Mecam,[390] ubi sepultus est Maoemet.[391]

Sciendum vero quod ventorum iniurias iugiter experti, diutius in salo [392] fluctuavimus. Tandem ad Cadiz, violencia venti compulsi, applicuimus,[393] fortissimum euri flatum excipientes.[394] Opidum autem incole deseruerant a tempore quo quidam Sarraceni in Silvia obsessi post casum urbis ad eos venerant, terrorem a nobis multiplicantes. Prefectus autem opidi ad nos exenia [395] deferens, supplicavit ut loco parceremus, promisitque sequenti die $\mathrm{XII}^{\mathrm{cim}}$ captivos Christianos se redditurum, insuper pecuniam quantamcumque congerere [396] potuisset. Statuta die tantum $\mathrm{IIII}^{\mathrm{or}}$. captivos adduxit, et visus est moras [397] pro aliis solvendis fraudulenter vindicare. Unde nostri eum indempnem

[384] Arzila, on the Atlantic coast of Spanish Morocco, some twenty miles southward from Cape Spartel, the northwestern extremity of Africa. Cf. Idrīsī, p. 202; U. S. Hydrographic Office, *East Atlantic Pilot*, p. 377.

[385] Tangier, ancient *Tingis*, the *Tanja* of the Arabs, some seven miles east of Cape Spartel, at the western entrance to the Strait of Gibraltar, the most important town in northern Morocco. Cf. Idrīsī, p. 201; U. S. Hydrographic Office, *Mediterranean Pilot*, I, 87–89; E. Lévi-Provençal, in *Encyclopaedia of Islam*, *s. v. Tangier*.

[386] Marrakech, the largest town in southern Morocco, one hundred and fifty miles south of Casablanca, which ultimately gave its name to the whole country. Cf. P. de Cénival, in *Encyclopaedia of Islam*, *s. v. Marrakesh*.

[387] MS. *astricti;* I follow Gazzera and Chroust in emending to read *Affrice*, but it seems barely possible that the manuscript reading is correct.

[388] MS. reading is clear; there is no warrant for either *Ogrimera* or *Agromera*, the readings respectively of Gazzera and Chroust. The land of the Ghomāra, an important Berber tribe of the western Maghrib, some of the clan names of which are still to be found at the present day among a number of Rif tribes. On the territory and the history of the tribe consult G. Yver, in *Encyclopaedia of Islam*, *s. v. Ghomāra*.

[389] The land of the Berbers, on whom consult R. Basset, in *Encyclopaedia of Islam*, *s. v. Berbers*. Barbaria, as used by the Christians in the twelfth and thirteenth centuries, seems to have applied generally to a good part of the whole north African littoral. Cf. H. C. Krueger, "The Routine of Commerce between Genoa and Northwest Africa during the Late Twelfth Century," in *Mariner's Mirror*, XIX (1933), 417–438.

[390] Mecca.

[391] Muḥammad.

[392] Chroust has quite unaccountably emended the text to read *salso* in place of *salo*.

[393] September 26.

[394] On the winds in the Strait of Gibraltar and its approaches see U. S. Hydrographic Office, *Mediterranean Pilot*, I, 38–41, noting especially the following statement apropos of the western approach: "Easterly winds prevail in summer, when they may last 15 days consecutively and blow hard all the time."

[395] Chroust, after Gazzera, emended to read *xenia*, but *exenia* is a not uncommon mediaeval form of the word.

[396] MS. *congeretere* or *congeregere*, a scribal corruption which in one letter is hardly legible.

[397] MS. *moris* was written first and then corrected.

abire fecerunt in festo Cosme et Damiani;[398] domos autem exusserunt, muros destruxerunt, vineas et ficulneas inciderunt, et illa die urbis destructioni, quantum potuerunt, insudavere.

Fuit autem Cadiz opidum valde oppulentum, a solis mercatoribus inhabitatum,[399] situm in insula[400] quam separabat a terra brachium maris Albee.[401] Insula autem parva est, sed ad aliam[402] stricta via terrestris est in mari eidem opido obnoxiam. Habuit autem opidum quinque munitiones, singillatim muris et turribus discretas, et amena valde habitacula; et ad ipsum confluere solebant ter in anno pro mercimoniis permutandis ex Affrica et Yspania Sarraceni, quia erat quasi in meditullio.[403]

Sequenti nocte[404] profecti sumus, et instanter velis innisi ventorum adversitate impediebamur;[405] et sicut solent naute, in diversa velificantes contrarietatem flatus arte delusimus, et sic in festo sancti [11v.] Michaelis[406] meridie strictum mare transsivimus. Sed quia maior pars classis adhuc contra ventos obluctabatur,[407] ad Ieziratarit'[408] accessimus, et anchoras figentes impugnare opidum proposuimus;[409] omnes etiam sequentes nos imitati sunt. In litore autem multos equites et pedites vidimus stare paratos defendere ripam; sed mulieres ad montana confugere. Nostri autem armati scaphas intraverunt. Sed quia non omnes unanimes[410] ad insultum se preparaverant, et precipue quia maxima tempestas orta est, tantum expectavimus tres naves que postreme[411] longam moram fecerant; et anchoris sublatis navigavimus

[398] September 27.

[399] MS. *inhabitaturum.*

[400] The island of León which is separated from the mainland by the Santi Petri Channel. Cf. U. S. Hydrographic Office, *East Atlantic Pilot*, pp. 207, 221.

[401] Apparently a corruption of some Arabic word, perhaps of *al-Baḥr*, the sea, referring to the Mediterranean which, according to most Arab geographers, began with the Gulf of Cádiz. Cf. C. F. Seybold, in *Encyclopaedia of Islam*, *s. v. Baḥr al-Maghrib.*

[402] Perhaps the small island of Santi Petri at the southern extremity of León Island. Cf. Idrīsī, p. 214; U. S. Hydrographic Office, *East Atlantic Pilot*, p. 221.

[403] This evidence of the prosperity and commercial importance of Cádiz in this epoch has been overlooked, the Muslim period being commonly represented as a time of decline and stagnation. Cf. C. F. Seybold in *Encyclopaedia of Islam*, *s. v. Cádiz;* U. S. Hydrographic Office, *East Atlantic Pilot*, pp. 207–221.

[404] September 28.

[405] Cf. above, p. 638, note 394.

[406] September 29. The date is confirmed by the *Gesta Regis Henrici Secundi*, ed. Stubbs, II, 90.

[407] MS. *obluctebatur* was written first and then corrected.

[408] MS. *Ieziratarit* of uncertain ending; Chroust read *Ieziratarif* without warrant in the manuscript. Tarifa; see above, p. 637, note 374.

[409] MS. *propsuimus.*

[410] MS. *hunanimes.*

[411] Chroust has emended to read *postremo.*

eiusdem diei crepusculo strictum mare exeuntes, ex utraque parte alta tantum montana conspicientes.

E regione Ierizitaris [412] ultra mare est Cac[ir] [413] Mucemuthe,[414] et inter hec castella est generalis transsitus de Affrica in Yspaniam et e converso. Habet autem strictum mare in latitudine duo nostra miliaria [415] et sex in longitudine, sicut perpendere potuimus.[416] A dextris in fine stricti maris reliquimus sequenti [die] [417] opulentissimam civitatem [418] Barbarie, ad quam confluunt omnes Christiani mercatores in Affricam [419] commercia transferentes, et maxime Ianuenses et Pisani hunc locum celebrant.[420] In eo etiam stacionarie sunt galee regis Marro*th*ie.[421] Item a leva in termino angusti [422] transsitus reliquimus Iezerita Hadra,[423] opidum bonum, et castellum Iebelatarie.[424]

Deinde spacioso [425] mari ad sinistrum latus nos committentes, prospero cursu has civitates transsivimus: Malagam,[426] Almonecam,[427]

[412] Evidently Tarifa, notwithstanding the variation of spelling. Cf. above, p. 637, line 2, p. 639, line 16, and notes.

[413] MS. *cac';* it is uncertain how the abbreviation should be extended, but it is evident that the author was endeavoring to reproduce the Arabic word Ḳaṣr as part of a place name. Gazzera and Chroust mistakenly read *castrum.*

[414] Al-Ḳaṣr al-Ṣaghīr (the *Ḳasr Masmūda* of Idrīsī and other Arab writers of his epoch), now in ruins, on the southern side of the Strait of Gibraltar, directly opposite Tarifa, beneath a spur of the Jebel Ghomāri, at the mouth of a small navigable river. Under the Almohades it was the principal African port of embarkation for Spain. It required a Portuguese fleet of eighty vessels and an army of 17,000 men to take it from the Muslims in 1458. See G. Yver, in *Encyclopaedia of Islam, s. v. al-Ḳaṣr al-Ṣaghīr;* cf. Idrīsī, p. 199; U. S. Hydrographic Office, *Mediterranean Pilot,* I, 91.

[415] That is, German or Teutonic miles. See above, p. 616 and note 86, p. 618 and note 113.

[416] For the true dimensions of the Strait see U. S. Hydrographic Office, *Mediterranean Pilot,* I, 60. Considering the probable length of the German mile, the author's estimate perhaps does not greatly understate the facts.

[417] September 30. It seems necessary to supply *die,* as Gazzera and Chroust have also done.

[418] Evidently Ceuta, opposite Gibraltar, on which see G. Yver, in *Encyclopaedia of Islam, s. v. Ceuta;* U. S. Hydrographic Office, *Mediterranean Pilot,* I, 93–95.

[419] MS. *Affricum.*

[420] On the traffic of the Genoese with this region see H. C. Krueger, "The Routine of Commerce between Genoa and Northwest Africa during the Late Twelfth Century," in *Mariner's Mirror,* XIX (1933), 417–438.

[421] MS. reading uncertain; possibly *Marrotzie,* the reading of Gazzera, is correct; Chroust read *Marrochie.* By *regis Marrothie* the author doubtless meant the Almohade caliph. Abū Yūsuf Ya'ḳūb al-Manṣūr was caliph 1184 to 1199.

[422] MS. apparently *angrati.*

[423] Algeciras, Arabic *al-Jazīra al-Khadrā',* "the green island," named after Isla Verde which lies in front of it. Cf. Idrīsī, pp. 199, 212–213; C. F. Seybold, in *Encyclopaedia of Islam, s. v. Algeziras;* U. S. Hydrographic Office, *Mediterranean Pilot,* I, 72–73.

[424] Gibraltar, on which see Idrīsī, p. 213; C. F. Seybold, in *Encyclopaedia of Islam, s. v. Gibraltar;* U. S. Hydrographic Office, *Mediterranean Pilot,* I, 71–83.

[425] MS. *spaciosa* was written first and then corrected.

[426] Malaga. Cf. Idrīsī, pp. 241, 250; U. S. Hydrographic Office, *Mediterranean Pilot,* I, pp. 101–103; E. Lévi-Provençal in *Encyclopaedia of Islam, s. v. Malaga.*

[427] Almuñécar, on the south coast of Spain some thirty-five miles eastward from Malaga. Cf. Idrīsī, pp. 242–243; U. S. Hydrographic Office, *Mediterranean Pilot,* I, 105.

Almariam,[428] Kartageniam,[429] Alacant,[430] Deniam,[431] Valenciam,[432] Burrianam,[433] Orpensam,[434] Pinnisculam.[435] Iuxta Betaieniam est Murcia.[436] Et sciendum quod hec via longissima est, quam vix fecimus in V. diebus et quinque noctibus fere continue et celerrime velificando. Sciendum est quod non vidimus [12r.] nisi altissimas rupes.

Tran[ssito mari primo occ]urrit [437] terra plana, et inde brevi spacio Ebora, Lebrus,[438] capacissimus et amplus in mare fluit, fluvius super quem sita est C*or*duba,[439] Turtasa,[440] versus montana a mari distans per

[428] Almería. Cf. Idrīsī, pp. 239–241, 245; C. F. Seybold, in *Encyclopaedia of Islam, s. v. Almeria;* U. S. Hydrographic Office, *Mediterranean Pilot,* I, 109–113.

[429] Cartagena. Cf. Idrīsī, p. 236; U. S. Hydrographic Office, *Mediterranean Pilot,* I, 124–125.

[430] Alicante. Cf. Idrīsī, p. 235; U. S. Hydrographic Office, *Mediterranean Pilot,* I, 132–134.

[431] Denia, province of Alicante, on the coast some fifty-five miles southeastward from Valencia. Cf. Idrīsī, pp. 233–234; C. F. Seybold, in *Encyclopaedia of Islam, s. v. Denia;* U. S. Hydrographic Office, *Mediterranean Pilot,* I, 139–140.

[432] Valencia. Cf. Idrīsī, pp. 232–233; E. Lévi-Provençal, in *Encyclopaedia of Islam, s. v. Valencia;* U. S. Hydrographic Office, *Mediterranean Pilot,* I, pp. 139–144.

[433] There is no warrant in the manuscript for *Buirianam,* the reading of Chroust. Burriana, province of Castellón de la Plana, near the coast some twenty-eight miles northeastward from Valencia. Cf. Idrīsī, p. 232; U. S. Hydrographic Office, *Mediterranean Pilot,* I, 145.

[434] Oropesa, province of Castellón de la Plana, on the coast at Cabo de Oropesa, some forty-two miles northeastward from Valencia. Cf. U. S. Hydrographic Office, *Mediterranean Pilot,* I, 146; *Enciclopedia Universal Ilustrada, s. v. Oropesa.*

[435] Peñíscola, province of Castellón de la Plana, on the coast some twenty miles northeastward from Cabo de Oropesa; according to Roger of Howden, writing of the year 1190 (*Chronica,* ed. Stubbs, III, 49), it was "ultimum castellum paganorum in Hispania supra mare." Cf. Idrīsī, p. 232; U. S. Hydrographic Office, *Mediterranean Pilot,* I, p. 148; *Enciclopedia Universal Ilustrada, s. v. Peñiscola.*

[436] Though the reading of the manuscript is certain, this sentence seems to present an insoluble difficulty. It would seem that *Murcia* can only be the important and well-known city of that name, and it would not be unnatural for the author to pause to mention it after his enumeration of the coastal points of interest to mariners. But all efforts to identify *Betaienia* have so far proved in vain. It is tempting to identify it with Beniajan, across the river and a little to the eastward of Murcia; but why should this inconspicuous village be used as a means of locating the capital of a province? Conceivably *Betaienia* is a corruption of *Wādi'l-abyaḍ* ("the white river"), an alternative Arabic name of the Segura, the river on which Murcia stands; but such a conjecture can hardly be entertained seriously. Kurth (p. 207, note 5), assuming that the sentence was "ganz verderbt," proposed to identify *Murcia,* not with the well-known city of that name, but with *Montsiá* (the *mons magnus qui dicitur Muncian,* or *Muscian,* of Roger of Howden, *Chronica,* ed. Stubbs, III, 47, 49), a conspicuous *sierra* extending along the coast southward from the Ebro to the little river Cenia, and to identify *Betaienia* with the latter stream; but such conjectures seem too extreme to be entertained.

[437] The text enclosed in brackets is wholly conjectural, a portion of the manuscript having decayed and been torn away.

[438] The River Ebro. There is no warrant in the manuscript for *Ebora vel Ebrus,* the reading of Chroust; *Lebrus,* as an alternative name, stands in apposition to *Ebora* and is set off between punctuation points, as it is printed between commas in the text above. It is, of course, possible that the author originally wrote *Ebora vel Ebrus* and that the present reading is due to a scribal error.

[439] MS. reading in part uncertain; *Corduba* is probably correct, but *Caduba,* the reading of Gazzera, is possible, as Chroust also recognized. It is not to be believed that the author intended to write *Corduba,* since he knew quite well that Cordova was on the Guadalquivir, and has so written (above, p. 635, line 12); the difficulty therefore seems to arise from some scribal error. The structure of the sentence seems to require that what the author originally wrote, or

duo nostra miliaria. Hec civitas prima Christianorum est quam Pisani et Ianuenses, tempore quo Ulixibona a nostris est capta,[441] ceperunt.[442] Inde incipit Catalonia,[443] terra optime[444] culta et castellis innumeris ornata. A Turtusa distat Terraconia[445] per unam dietam, civitas olim maxima sed nunc parva, in qua sedes est archiepiscopalis magne dignitatis. Inde ad dietam est Barcelona,[446] ubi est caput[447] comitatus A*r*talonensis,[448] ab inde ad sex dietas Narbona,[449] inde ad II. Monspessulanus,[450] inde ad III. Massilia.[451] Et notandum quod postea Massilie et in Montepessulano vidimus mercatores qui in civitatibus Sarracenorum erant cum transsivimus; et nos viderunt et dixerunt quod omnes Sarraceni ita pavefacti erant de transsitu nostro quod nullam civitatem defendissent si eam adiissemus, sed tantum ad fugam se preparabant.[452]

intended to write, was some alternative form for the name of Tortosa (as in the case of *Ebora* and *Lebrus* above), but what that form was now seems beyond determination. Kurth (p. 207, note 6) proposed to identify *Caduba*, as he believed the form of the name to be, with Isla de Buda, a part of the delta between two branches of the Ebro at its mouth. Cf. U. S. Hydrographic Office, *Mediterranean Pilot*, I, p. 152.

[440] Tortosa, on the left bank of the Ebro, above the delta, "against the mountains," as the author says. Cf. E. Lévi-Provençal, in *Encyclopaedia of Islam, s. v. Tortosa.*

[441] The reference is to the Christian conquest of Lisbon in 1147, on which see *De Expugnatione Lyxbonensi*, ed. David. The date indicated is erroneous by more than a year.

[442] Tortosa was taken after a six months' siege on December 30, 1148, by Ramon Berenguer IV, count of Barcelona, with the assistance of the Genoese and other allies, among whom the Pisans seem not to have been included. Cf. Caffaro, *Annales Genuenses*, in *M.G.H.*, *Scriptores*, XVIII, 21, 38–39; Jerónimo Zurita, *Anales de la Corona de Aragon* (Saragossa, 1610), I, 62–65; A. de Bofarull y Brocá, *Historia Crítica de Cataluña* (Barcelona, 1876–78), III, 31–36; Flórez, *España Sagrada*, XLII, 108–113; E. Lévi-Provençal, in *Encyclopaedia of Islam, s. v. Tortosa,* and the Arabic sources there cited.

[443] The statement is very nearly correct, the modern boundary between Catalonia and Valencia being the little river Cenia a few miles to the south of the Ebro.

[444] MS. *obtime.*

[445] Tarragona, ancient *Tarraco,* capital of the modern province of Tarragona, some thirty miles northeastward from the mouth of the Ebro; it is still an archiepiscopal see. Cf. E. Lévi-Provençal in *Encyclopaedia of Islam, s. v. Tarragona.*

[446] Barcelona, on which cf. C. F. Seybold, in *Encyclopaedia of Islam, s. v. Barcelona.*

[447] MS. *capud.*

[448] MS. possibly *Aitalonensis*, as Chroust read. Chroust, following Gazzera, emended the text to read *Catalonensis* but *Barcelonensis* would be more in accord with twelfth century usage since Ramon Berenguer IV and his son Alfonso II of Aragon commonly bore the title of "count of Barcelona." Cf. *Coleccion de Documentos Inéditos del Archivo General de la Corona de Aragón*, IV, No. CLXVIII.

[449] Narbonne, department of Aude, in the south of France.

[450] Montpellier, department of Hérault, in the south of France. It connects with the Mediterranean by the River Lez at Grau de Palavas.

[451] Marseilles.

[452] MS. *properabant.*

Appendix A

Silves: Its Situation, Fortifications, and History Under Muslim Rule[1]

Silves enjoys an extraordinary situation which under the conditions of mediaeval warfare was one of great strategic importance. It stands on an isolated hill which rises to a height of almost two hundred feet above the right bank of the little river Arade. It is joined to the left bank by a bridge of low arches which is believed to be of Moorish origin,[2] and it was undoubtedly connected with the coast in the Middle Ages by a direct road over land as well as by the river route.[3] Though a small port at the head of tidewater and only some eight miles from the sea, it gives the impression of being an inland town, so completely does it appear to be cut off from the outside world by the rugged hills which hem in the river valley.

On its northeastern slope the hill of Silves is so precipitous as to be practically bare except for its fortifications and to this day there is no entrance to the city from this side; but on the southwest it descends more gently towards the riverbank and to the flat low-lying land along the river to the west. The modern town—of somewhat less than 10,000 inhabitants—climbs, like its mediaeval predecessor, up the southwestern slope from the river towards the summit of the hill; but it does not spread out to any great extent over the low flat land at its base, and there can be little doubt that the modern population is considerably smaller than the mediaeval.

The mediaeval fortifications of Silves, when fully developed,[4] must have presented on a large scale a striking example of the kind of defensive works which were typical of mediaeval Spain, whether Muslim or

[1] Unfortunately the imposing remains of the fortifications of Silves have never been made the subject of a thorough archæological investigation, and their early history is shrouded in darkness. There appears to be no accurate survey or map of the modern town and ruins; and though the fortress has been officially declared a *Monumento Nacional*, not even air photographs seem to be obtainable. What follows, therefore, is a mere sketch, based mainly on direct but hurried observation. Worthy of consultation is Pedro P. Mascarenhas Júdice, *A Sé e o Castelo de Silves* (Gaia, Portugal, 1934), which contains a bibliography and a number of helpful photographs The plan of Silves in Silva Lopes, *Relação*, opposite p. 108, is inaccurate and of little use.

[2] Júdice, *op. cit.*, p. 58, and photograph opposite p. 40.

[3] See text above, pp. 617–618.

[4] It is, of course, not certain that they were fully developed at the time of the siege by the Portuguese and the crusaders in 1189, but one gains the impression that they were already very extensive. The city had already undergone one important siege and been taken in 1063. See below, p. 650.

Christian, and which have been described by a recent authority as follows:

Le château, qu'il fût entre les mains des Musulmans ou des Chrétiens, présentait presque toujours la même disposition générale. Il se dressait sur un piton; parfois, du côté de la pente la plus raide et la moins accessible aux assiégeants, il n'avait pas de mur de défense. Mais en général, un rempart en faisait tout le tour: il était construit en pierre ou en béton et flanqué de tours et de bastions, d'où les assiégés essayaient de repousser les attaques. Cette enceinte au sommet du piton était, dans la plupart des cas, fort peu spacieuse et formait le réduit (*ḥaram al-ḥiṣn*); au-dessous d'elle, il y avait une plateforme (*sāḥa*) épousant la forme du terrain, et sur laquelle s'ouvrait l'agglomération groupée autour du château fort, le faubourg ou *rabad*. C'est là que logeaient les garnisaires avec leurs familles et les petits commerçants de l'endroit.[5]

At Silves there was a strong rampart (*almadina*) girdling the entire hilltop, and it evidently enclosed a considerably larger area than was common in strongholds of lesser importance. From it, more or less parallel walls, or so it seems, descended to the riverbank and to the source of fresh water supply (*conductus aquarum*), which latter in turn was especially protected by a fortress of four strong towers (*corrasce*). And the suburb (*rovalle*) which spread out over the flat land at the base of the hill was also enclosed within a wall, though not a very strong one, and further protected by a watch tower (*alverrana*).[6]

Unfortunately practically all of the lower fortifications of mediaeval Silves have disappeared without leaving a trace above ground, and any description of the existing remains of the stronghold must be confined to the ramparts higher up. These, however, are imposing enough. Generally speaking they encircle the hill, except on the southwest where they have been destroyed by the development of the modern city,[7] and they are high and strong. There are still some seventeen towers [8] in a fair degree of preservation and much of the intervening wall still stands between them. Both towers and walls are of that peculiar "mud" or concrete construction which the Portuguese call *táipa* and the Spanish, *tapia* or *hormazo* and which was already described by Pliny as being

[5] E. Lévi-Provençal, *L'Espagne Musulmane au Xème Siècle*, pp. 150–151. The author goes on to give an example of such a *château fort* in Catalonia which was taken by the Muslims in 1003 and carefully preserved from destruction and immediately occupied by a Muslim population and garrison who continued the existing arrangements. Note that the author of the text above (p. 621), after describing the fortifications of Silves with the use of a number of words which were strange to him, continues: "Et nota quod hec nomina sunt appellativa, non propria. Ubicumque enim tales dispositiones sunt civitatis in terra illa, tam a Christianis quam a paganis, talia habent nomina."

[6] See text above, pp. 619–621, *et passim;* cf. Ralph de Diceto, *Opera Historica*, ed. Stubbs, II, 66.

[7] Possibly some of the destruction has been due to earthquakes.

[8] "Munita turribus, sicut dicitur, circiter quadraginta," says Ralph de Diceto (*Opera Historica*, ed. Stubbs, II, 66); but he wrote from hearsay and not improbably exaggerated; yet it is certain that there were a number of towers in the Middle Ages which have since disappeared.

characteristic of both Africa and Spain;[9] and though they were once faced with stone masonry, much of the facing has since fallen away, or been pulled down for use in modern buildings, leaving exposed surfaces which enable the modern observer to draw some conclusions concerning the original method of construction.[10] Evidently the *táipa* was tamped in between planks of wood in courses some thirty-four inches thick; and it would seem that the planks were prevented from spreading apart by means of ropes which were cut when the planks were moved upward to serve as a frame for the next higher course. Holes are still visible in the walls where the ropes have rotted away, and they are in some cases large enough to permit a man to thrust his arm into them for a considerable distance.

A notable feature of most of the towers is that they stand outside the wall at a distance of some eight or ten feet and are unconnected with it except by high round arches which support passageways into them from the top of the wall. A number of the towers measure about sixteen feet square, but some are larger; and there is one great tower in particular, standing beneath the city gate (*Porta da Cidade*), which is some forty feet square. The towers are, as a rule, solid (that is, they contain no chamber) from the ground up to a height which is even with the top of the wall; but at this level each contains a square room, above which there was evidently a battlemented fighting platform.

Our best mediaeval description of Silves, that of the twelfth-century Arab geographer Idrīsī, is so highly informing as to merit full quotation:

> Silves, jolie ville bâtie dans une plane, est entourée d'une forte muraille. Ses environs sont plantés en jardins et en vergers; on y boit l'eau d'une rivière qui baigne la ville du côté du midi, et qui fait tourner des moulins. La mer Océane en est à trois milles du côté de l'occident. Elle a un port sur la rivière et des chantiers. Les montagnes environnantes produisent une quantité considérable de bois qu'on exporte au loin. La ville est jolie et l'on y voit d'élégants édifices et marchés bien fournis. Sa population ainsi que celle des villages environnants se compose d'Arabes du Yémen et d'autres, qui parlent un dialecte arabe très-pur; ils savent aussi improviser des vers, et ils sont tous éloquents et spirituels, les gens du peuple aussi bien que les personnes des classes élevées. Les habitants des campagnes de ce pays sont extrêmement généreux; nul ne l'emporte sur eux sous ce rapport. La ville de Silves fait

[9] *H. N.*, xxxv, 169: "Quid? non in Africa Hispaniaque e terra parietes quos appellant formaceos, quoniam in forma circumdatis duabus utrimque tabulis inferciuntur verius quam struuntur, aevis durant, incorrupti imbribus, ventis, ignibus, omnique caemento firmiores?" The passage is repeated almost verbatim by Isidore of Seville (*Etymologiae*, xv, 9, 5), who calls such construction *formatum* or *formacium*.

[10] *Mouros de táipa* are still builded in the Algarve. The method of construction has been described with some fulness by Silva Lopes, *Relação*, pp. 74–75.

partie de la province d'Ach-Chinchîn [*al-Shinshīn*], dont la territoire est renommé par ses jardins plantés de figuiers; on exporte ces figues vers tous les pays de l'Occident; elles sont bonnes, délicates, appétissantes, exquises.[11]

Silves [12] and its environs, lying as they did in the far southwestern corner of the Iberian Peninsula, received but scant attention from Arabic writers until a comparatively late date. The circle, or *kurā*, which Idrīsī calls al-Shinshīn, and which according to Abu'l-Fidā', doubtless writing of a later period, was named after Silves itself,[13] had in an earlier age been known among the Muslims as Ocsonoba,[14] from the ancient *Ossonoba* [15] situated some distance to the east of Silves. Presumably the region fell into the hands of Mūsā b. Nuṣair, the Arab conqueror of Spain, in 712 without any serious resistance; and it may well be that it received the predominantly Yemenite population of which Idrīsī speaks at the time of the first settlement. However, there is some evidence that the district also received an infusion of North Arabian or Syrian (Kaisite) settlers from among the companions and followers of Balj b. Bisher in 743; [16] and it is not to be believed that the earlier Spanish population which accepted Islam and learned Arabic was wholly eliminated.[17]

Indeed, when in the late ninth century we first get any information concerning the political organization of this region, we find it under the control of a neo-Muslim or renegade family. According to Dozy, in the closing years of the reign of Muḥammad I, at a time when the Cordovan emirate had quite generally lost control over the local lords or nobles, Yaḥyā b. Bakr, whose grandfather had been a Christian

[11] Idrīsī, p. 217 (translation of Dozy and de Goeje).

[12] The historical matter which follows has unfortunately had to be assembled without command of the Arabic language. It is hoped, however, that little of importance has been missed.

[13] *Géographie d'Aboulféda* (French translation by [J. T.] Reinaud, Paris, 1848–83), II, 237: "c'est à la fois le nom d'un canton et d'une ville."

[14] Cf. E. Lévi-Provençal, in *Encyc. of Islam*, *s. v. Ocsonoba*, and the references there given. Ibn 'Idhārī, *al-Bayān al-Mughrib* (French translation by E. Fagnan, Algiers, 1901–04), II, 226, in speaking of Santa Maria de Algarve (modern Faro) locates it in "the canton of Ocsonoba."

[15] *Ossonoba* has been identified with the site of modern Milreu near Estoy, a few miles to the north of Faro; but it may well be questioned whether the identification should not rather be with Faro itself (cf. E. Lévi-Provençal, in *Encyc. of Islam*, *s. v. Ocsonoba*). As indicated above, note 14, Ibn 'Idhārī in one passage applies the name to a canton and locates Faro within it; but in another passage (*op. cit.*, II, 332) in which he refers to the town (*ville*) of Ocsonoba, it is difficult to believe that he is not really speaking of Faro. Ibn al-Abbār (*Scriptorum Arabum Loci de Abbadidis*, ed. R. Dozy, Leyden, 1846–52, II, 123; Portuguese translation in *Revue Hispanique*, IX, 1902, p. 61) calls Faro "the port of Ocsonoba."

[16] See R. Dozy, *Histoire des Musulmans d'Espagne*, new ed. by E. Lévi-Provençal (Leyden, 1932), I, 168–169, and the references there cited.

[17] On the whole subject of the settlement of Spain after the Muslim conquest and of the distribution of the rival ethnic groups, see E. Lévi-Provençal, *L'Espagne Musulmane au X^ème^ Siècle*, pp. 8–28.

named Zadulpho, had gained control first of Santa Maria de Algarve (modern Faro) and then of the whole province of Ocsonoba [18] and had declared his independence.[19] Ibn ʿIdhārī writes almost with enthusiasm of the prosperous rule of Yaḥyā's son and successor, Bakr, during the reign of ʿAbd-Allāh (889–912). He converted Santa Maria into a stronghold and provided it with iron gates. He gathered abundant supplies and had an army of brave and well equipped soldiers. He had a fully developed administration, with a council and a department of finance. He was united in firm alliances with the lords of Badajoz and Beja and Mértola and esteemed his power to be comparable with that of the lord of Seville. According to Dozy, the emir of Cordova had offered him the governorship of the province and he had accepted the offer without recognizing that it imposed any serious obligation upon him.[20] Far from considering travelers and merchants as victims to be plundered, he commanded his subjects to protect them and grant them hospitality, and his orders were so well obeyed that the traveler in the province of Ocsonoba found himself everywhere among friends and, as it were, among kinsmen. One can only conclude from the account of Ibn ʿIdhārī, which we have been following, that Santa Maria was the capital or centre of Bakr's administration, but Dozy declares that Bakr himself resided at Silves.[21]

The power and prosperity of this petty renegade lordship extended a good way into the tenth century and into the reign of the greatest of the Umaiyads of Spain. Bakr was evidently succeeded by his son Khalaf b. Bakr who ruled without regard to the emirs of Cordova, and paid no tribute, until 929 when the triumphant ʿAbd al-Raḥmān III, having taken Beja, came to camp before his walls. Khalaf evidently recognized that armed resistance would be useless, but he managed to avoid a complete overthrow. He sent envoys to the emir to make offers of friendship and submission and to explain that his failure to pay tribute had been due to the remoteness of the land in which he dwelt. He also sent rich gifts and offered full payment of a handsome tribute for the future. And since the inhabitants of the country manifested a warm attachment to him and praised his administration, ʿAbd al-Raḥmān for once departed from his usual custom and continued Khalaf in his

[18] It can hardly be doubted that Silves would have been included in this territory.

[19] *Histoire des Musulmans d'Espagne*, new ed., II, 57. Dozy does not give the source of his information at this point, and I have failed to discover it.

[20] *Ibid.* Again I have unfortunately failed to identify the source of Dozy's information.

[21] Most of the matter in the foregoing paragraph comes from Ibn ʿIdhārī (*al-Bayān al-Mughrib*, French translation by E. Fagnan, II, 223, 226) who was used by Dozy (*Histoire des Musulmans d'Espagne*, new ed., II, 57), but it seems evident that Dozy had at his disposal some other source of information which I have failed to discover.

post, consenting that he should become, not his subject, but his vassal, his tributary, on condition that he pay regular annual tribute and harbor no rebels or fugitives.[22]

Thus, if we can rely upon Ibn 'Idhārī, it appears that in the second half of the ninth century and the first half of the tenth Santa Maria (Faro) rather than Silves was the chief town and administrative centre of the little province which has survived as the Algarve in modern Portugal.

The earliest direct mention of Silves which I have noted in the sources refers to the year 844 or to a time very shortly thereafter. After the devastating descent of the Northmen (*al-Majūs*) on Muslim Spain (al-Andalus) in that year, 'Abd al-Rahmān II, the emir of Cordova, appears to have entered into diplomatic relations with the Majūs "king"; and we are told that his ambassador, al-Ghazzāl, embarked at Silves when he set out upon his journey to the northern court.[23] Silves is again mentioned in connection with Northmen in 966 when its port was the scene of an important naval victory of the Muslims over the northern raiders,[24] but what its development as a city or centre of administration had by that time been we have no means of knowing.

The history of Silves emerges into a clearer light in the eleventh century when for a time it became the capital of one of the petty dynasties of so-called Taifa kings (Arabic, *Mulūk al-Ṭawā'if; Spanish, Reyes de Taifas*) which arose from the ruins of the fallen Umaiyad caliphate of Cordova; and fortunately the recent publication of a new Arabic source [25] has made it possible to correct the older accounts and

[22] Ibn 'Idhārī, *al-Bayān al-Mughrib*, French translation by E. Fagnan, II, 332–333; cf. E. Lévi-Provençal, *L'Espagne Musulmane au Xème Siècle*, pp. 120–121. Ibn 'Idhārī here speaks as if Ocsonoba were a town, Khalaf's capital, before the walls of which 'Abd al-Raḥmān encamped; but we know of no walls at *Ossonoba* (if it is to be identified with Milreu: see p. 646, note 15, above), whereas we do know that Bakr had fortified Santa Maria. The soundest conculsion would, therefore, seem to be that Ibn 'Idhārī has fallen into some confusion and that Santa Maria (Faro) was the centre of Khalaf's power. Dozy (*Histoire des Musulmans d'Espagne*, new ed., II, 111–112) adds nothing to the account of Ibn 'Idhārī at this point, and he makes no further mention of Silves as a residence of the renegade lord of Ocsonoba.

[23] See the account of the mission, first published and translated by R. Dozy in his *Recherches sur l'Histoire et la Littérature de l'Espagne pendant le Moyen Âge*, II, 269–278, and Appendix pp. lxxxi-lxxxviii.

[24] Ibn 'Idhārī, *al-Bayān al-Mughrib*, French translation by R. Dozy, *op. cit.*, II, 289: "Ensuite la flotte musulmane sortit de la rade de Séville et attaqua celle des Madjous dans la rivière de Silvès. Les nôtres mirent plusieurs vaisseaux ennemis hors de combat, délivrèrent les prisonniers musulmans qui s'y trouvaient, tuèrent un grand nombre de mécréants et mirent les autres en fuite."

[25] An historical fragment on the *Mulūk al-ṭawā'if*, published by E. Lévi-Provençal in his edition of the third volume (completing Dozy's edition of the earlier part) of Ibn 'Idhārī, *al-Bayān al Mughrib* (Paris, 1930), and translated by him as an appendix in his new edition of R. Dozy, *Histoire des Musulmans d'Espagne*, III, 215–235; cf. *idem*, in *Encyc. of Islam, s. v. Silves*. The editor, whose competence is beyond question, accepts the fragment as a trustworthy authority.

write the history of the little dynasty of Silves, that of the Banū Muẓain, with some certainty and definiteness.

The district of Silves seems to have maintained its loyalty to Cordova longer than most of the local divisions of Muslim Spain; [26] but in the year of the Hijra 440 (A. D. 1048–49) it gained its independence under the leadership of its ḳāḍī Abu'l-Aṣbagh 'Isā b. Abī Bakr, an Arab official who traced his descent from one Muẓain b. Mūsā, the first of his ancestors who had come to Spain [27]—presumably at the time of the Muslim conquest and settlement in the first half of the eighth century [28] —and who signalized his new authority by the assumption of the honorific title of *al-Muẓaffar.* He was a man of energy with a will to rule, who had taken advantage of a period of disorder and rebellion to raise himself to supreme power: and, once installed in his new position, he had known how to make himself popular through a distribution of wealth. He strengthened the defenses of the city, put his army upon a wartime footing and took careful precautions against Abū 'Amr 'Abbād, called *al-Mu'taḍid,* the 'Abbādid ruler of Seville and most powerful of all the Taifa kings of Spain. He began by sending presents and seeking al-Mu'taḍid's friendship, but since all his efforts in this direction proved in vain and the king of Seville launched repeated incursions against him, he placed himself at the head of his forces and resorted to war. But the war was unsuccessful. After numerous defeats al-Muẓaffar was finally deposed by his more powerful rival and put to death in the spring of 1054 after a turbulent reign of five years.[29]

Yet, strange to say, the Taifa dynasty of the Banū Muzain did not end with the overthrow of its founder, and Silves was not yet to become

[26] Cf. Dozy, *Histoire des Musulmans d'Espagne,* new ed., II, 308.

[27] *Ibid.:* III, 220: "'Isā b. Abī Bakr Muḥammad b. Sa'īd b. Ǧamīl b. Sa'īd—l'auteur du commentaire *d'al-Muwaṭṭā'* [of Mālik b. Anas]—b. Ibrāhīm b. Abī Naṣr Muḥammad b. Ibrāhīm b. Abī'l-Ǧawd Muzain b. Mūsā (Muzain était celui de ses ancêtres qui vint s'installer dans al-Andalūs), seigneur de Šilb (Silves)." According to Dozy (*Ibid.,* p. 53), citing Ibn al-Abbār, his ancestors had possessed extensive properties in this part of the Iberian Peninsula and had often held important posts in the public service in the time of the Umaiyads.

[28] Cf. E. Lévi-Provençal, *L'Espagne Musulmane au Xème Siècle,* p. 20.

[29] Dozy, *Histoire des Musulmans d'Espagne,* new ed., III, 221: "C'était un homme qui montrait son autorité et son énergie dans les jugements qu'il rendait et dans tous ses actes. Quand il vit apparaître la période de troubles, il se mit dans la ville en état de révolte. Les habitants de Silves et de toutes les régions voisines le proclamèrent, en l'année 440. . . . Quand son autorité fut rétablie, il organisa solidement la défense de la ville, en rassembla les hommes et en partagea entre eux les richesses; il en mit sur pied les troupes, se tint avec le plus grand soin en garde contre al-Mu'tadid et commença à lui faire des présents et à lui demander la paix. Mais cela ne lui servit de rien: al-Mu'tadid lançait contre lui des incursions chaque jour, à chaque moment. Voyant qu'il ne pouvait mettre fin à son hostilité, malgré les amabilités qu'il lui prodiguait, al-Mudaffar se lança en personne contre lui à la tête de ses troupes et de ses partisans. Des combats et des engagements eurent lieu entre les deux partis, avec de nombreuses pertes; le succès revenait toujours à al-Mu'taḍid qui finit par déposer et tuer son adversaire, à la fin de 445."

a mere dependency of Seville. One can only conjecture that for the time being al-Mu'tadid was too heavily engaged in other parts of the Peninsula to follow up his victory in the far Southwest. Whatever the explanation, our source states specifically that in accordance with al-Muẓaffar's will and on the very day of his death, his cultivated son Abū 'Abd Allāh Muḥammad b. 'Isā, called *al-Nāṣir,* was proclaimed his successor and received the full allegiance of his subjects; and in some way he managed to avoid the hostility of the king of Seville and reigned prosperously and in peace until his death towards the middle of 1058.[30]

Once more the throne passed peacefully by testament, this time to al-Nāsir's son, 'Isā b. Muḥammad, who assumed his grandfather's honorary title, *al-Muẓaffar,* and set out to rule in accordance with the wise policies of his father. But he was not so fortunate as to be allowed to reign in peace. Al-Mu'tadid renewed the war which had been suspended during al-Nāṣir's reign; and in the end Silves was besieged and taken by him after its walls had been broken by siege engines on one side and undermined by sappers on the other. Al-Muẓaffar was seized in his palace, doubtless the palace of al-Sharājib,[31] and beheaded (Sept.–Oct., 1063), and the little dynasty of the Banū Muẓain and the independence of Silves were at an end, after an existence of only fifteen years.[32]

Thereafter, if we can trust the somewhat romantic history which 'Abd al-Wāḥid al-Marrākushī has woven around the name of the ambitious and talented poet-politician of Silves, Abū Bakr Muḥammad Ibn 'Ammār,[33] al-Mu'taḍid placed the administration of newly conquered

[30] *Ibid.:* "Il fut proclamé conformement au testament de son père, le jour de la mort de ce dernier, dans le district de Silves, à la fin de l'année 445. . . . Il prit le nom d'an-Nāṣir; du vivant de son père, il avait adopté celui de 'Ubaid ad-dawla. Il reçut complètement le serment d'allégeance: on l'aimait pour son goût pour l'étude, sa culture et l'ampleur de ses connaissances. Il demeura prince de Silves jusqu'à sa mort, qui survint dans cette ville en rabī' II 450."

[31] The name comes from al-Mu'tamid's poem on Silves cited below, p. 651.

[32] Dozy, *Histoire des Musulmans d'Espagne,* new ed., III, 221–222: "Il fut proclamé le jour de la mort de son père, en rabī' II 450. . . . La même région qui avait prêté serment à son père, le lui renouvela. Il suivit la conduite de son père jusqu'au moment où al-Mu'taḍid lui chercha querelle: il fit faire contre lui des coups de main et lança contre lui des pointes de cavalerie. Puis il vint l'assiéger étroitement et lui coupa tous les approvisionnements en charbon et en bois. La situation devint mauvaise et empira pour les habitants de Silves et du reste du pays et al-Mu'taḍid enleva en fin de compte de vive force la ville à Ibn Muzain, après avoir défoncé d'un côté le rempart au moyen de machines de guerre et l'avoir sapé d'un autre. Il pénétra dans son palais, le captura et lui trancha la tête, dans son injustice contre lui et son audace envers Allāh Grand et Puissant. Cela se passait en šawwāl 455. . . . Son règne avait été de cinq ans. La dynastie des Banū Muzain s'éteignit et leur royauté prit fin. La durée est le privilège d'Allāh très Haut!" Santa Maria (Faro) was not subject to the Banū Muẓain of Silves, but was independent under its own Taifa dynasty until conquered by Seville: *ibid.,* p. 222. Dozy's statement (*ibid.,* p. 54) that al-Mu'taḍid joined it to Silves, if true, must refer to 1063 or later.

[33] *Histoire des Almohades,* French translation by E. Fagnan, pp. 94–110. Regarding his origin and education the author says: "[Il] était originaire d'une bourgade nommée Chenna-

Silves and its dependent territory under his son al-Muʿtamid.[34] The latter, who was also a poet, associated Ibn ʿAmmār with himself in the government of Silves, making him his vizier, and the intimacy between them was so close that on one occasion after a *soirée littéraire* in the prince's palace they shared the same pillow. The delights of their life in Silves left an indelible impression upon the emotions and the memories of both the young men, but the rumors which reached al-Muʿtaḍid's ears concerning their association together led him to banish Ibn ʿAmmār from the realm.[35] However, when al-Muʿtamid succeeded his father on the throne of Seville in 1069, he promptly recalled his friend from exile [36] and elevated him to a position of highest favor. At his own request Ibn ʿAmmār was made governor of Silves and its dependent territory with full powers over both domestic and foreign affairs, and he continued to hold that office until al-Muʿtamid, consumed with a desire to see him and unable longer to endure his absence, recalled him to Seville and made him vizier of the kingdom.[37] On the occasion of Ibn ʿAmmār's departure to take up his post in Silves, al-Muʿtamid addressed him in a poem, reminiscent of their joyous and carefree life there together, which is too well-known to warrant reproduction here.[38] Our last glimpse of Silves in this epoch is contained in ʿAbd al-Wāḥid al-Marrākushī's description of the new governor's entry into the city, which may be quoted in full: "Ibn ʿAmmâr fit son entrée entouré d'un imposant cortège et d'une foule d'esclaves et de courtisans, en déployant

boûs, faisant partie du territoire de Silves et où ses ancêtres étaient aussi établis. Sa famille était obscure, et ni lui ni aucun de ses ascendants n'avait, à aucune époque, exercé de fonctions administratives, ou tout au moins cela n'est dit d'aucun membre de cette famille. Il alla tout jeune à Silves, où il fut élevé et où il étudia les belles-lettres, sous la direction de plusieurs maîtres, et entre autres d'Aboû'l-H'addjâdj Yoûsof b. ʿIsa el-Aʿlem. De là il se rendit à Cordoue, où il continua les mêmes études et où il devint fort habile en poésie, si bien qu'il fit de ce talent son gagnepain." Dozy (*op. cit.*, III, chs. 9–11) has devoted much space to him.

[34] Dozy (*op. cit.*, III, 54) has dated the prince's rule of Silves in 1052, but if we are to accept the authority of the historical fragment which we have been following, it must be dated in 1063 or shortly thereafter: the fact that al-Nāṣir followed his father immediately on the throne of Silves precludes the possibility of its being dated in 1054. ʿAbd al-Wāḥid al-Marrākushī (*Histoire des Almohades*, French translation by E. Fagnan, p. 99) is indefinite as to date.

[35] ʿAbd al-Wāḥid al-Marrākushī, *Histoire des Almohades*, French translation by E. Fagnan, p. 99.

[36] *Ibid.*, p. 97: "[Ibn ʿAmmār] fut nommé par El-Moʿtamid . . . , à peine monté sur le trône, au gouvernement de la ville de Silves et du territoire en dépendant."

[37] *Ibid.*, p. 101: "A l'avénement de Moʿtamid, Ibn Ammâr lui demanda le gouvernement de Silves, d'où, nous l'avons dit, il était originaire et où il avait été élvé. Le prince le lui accorda avec les pleins pouvoirs les plus étendus, tant pour les affaires intérieures que pour les extérieures, et il l'exerça jusqu'au jour où Moʿtamid, dévoré par le désir de le revoir et incapable de supporter plus longtemps son absence, lui retira ce poste et le rappela auprès de lui en qualité de vizir."

[38] The original in *Scriptorum Arabum Loci de Abbadidis*, ed. R. Dozy, I, 39, 84; French translation by Dozy in his *Histoire des Musulmans d'Espagne*, new ed., III, 91; English translation in the English edition of Dozy's work (London, 1913), pp. 668–669.

plus de faste que n'avait fait Mo'tamid lorsqu'il gouvernait cette ville du vivant de son père Mo'tad'id." [39]

Presumably Silves remained in subjection to Seville for the remainder of the 'Abbādid period and passed with it under the domination of the Almoravides in 1091, or shortly thereafter.[40] Our sources remain silent concerning it until it emerges into prominence in connection with the revolt of the Spanish Muslims of the Southwest (*al-Gharb al-Andalus*) against the rule of the Almoravides in the period of their decline towards the middle of the twelfth century.[41]

The central figure in the revolt of al-Gharb was a native of Silves named Ibn Ḳasī,[42] an ambitious but not very capable man of Christian or neo-Muslim (renegade) origin, who had passed a part of his youth in his native city and is said to have been employed in its financial administration.[43] He first attracted public notice as a religious leader. Having traveled in Spain and Morocco and entered into relations with the chief of the *Ṣūfī* mystics of Almería, and having returned to his own country and expounded in public the books of the great Muslim philosopher-theologian al-Ghazālī,[44] he managed to become the acknowledged leader of Sufism in al-Gharb. He seems also to have given some new turn to the sect and to have conferred on his followes the special title of *Morīdīn* (Adepts); hence the revolt in which he played so prominent a part is sometimes called "The Rebellion of the *Morīdīn.*" [45] The new sect met with considerable success in al-Gharb, principally in Silves, Niebla, and the region of Mértola. It had a kind of monastery at Arrifana on the desolate west coast to the north of Cape St. Vincent [46] where the *Morīdīn* gathered in considerable numbers, and from which,

[39] *Histoire des Almohades*, French translation by E. Fagnan, pp. 97–98.

[40] On the conquest of al-Andalus by the Almoravides see Ibn Khaldūn, *Histoire des Berbères*, ed., Casanova, II, 76–78; R. Menéndez Pidal, *La España del Cid*, pts. iv–v; Francisco Codera, *Decadencia y Desaparición de los Almoravides en España*, pp. 1–28; Dozy, *Histoire des Musulmans d'Espagne*, new ed., III, chs. 12–14.

[41] Much of what follows concerning the revolt is derived from the careful study of Codera, *op. cit.*, pp. 29–52. The principal sources are Ibn al-Abbār, *al-Ḥulla al-Siyarā'*, ed. R. Dozy, in his *Notice sur Quelques Manuscrits Arabes* (Leyden, 1847–51) pp. 199–205, 239; 'Abd al-Wāḥid al-Marrākushī, *Histoire des Almohades*, French translation by E. Fagnan, p. 122; Ibn Khaldūn, *Histoire des Berbères*, ed. Casanova, II, 183–188.

[42] Abu'l-Ḳāsim Ahmed b. al-Ḥusain Ibn Ḳasī, commonly called *Abencasi* by Spanish writers. Cf. E. Lévi-Provençal, in *Encyc. of Islam*, *s. v. Silves.*

[43] Ibn al-Abbār, in Dozy, *Notice sur Quelques Manuscrits Arabes*, p. 199; Codera, *op. cit.*, p. 33.

[44] Ibn al-Abbār, *loc. cit.*

[45] Codera, *op. cit.*, pp. 33–34.

[46] For identification of the site with Arrifana see C. F. Seybold, "Monchique et Arrifana d'Algarve chez les Auteurs Arabes," in *O Archeologo Português*, VIII (1903), 125–126; cf. Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 202. Arrifana is situated near Aljezur some twenty-five miles northward of Cape St. Vincent. See *Guia de Portugal*, II, 320, and map opposite p. 260. Also see map below, facing p. 666.

according to one authority, reports of spurious miracles of Ibn Ḳasī were spread abroad.[47]

In the spring of 1144 (Shawwāl, 538 A.H.), no doubt after there had been much revolutionary as well as religious propaganda, one of Ibn Ḳasī's followers began a revolt by seizing a castle near Silves called *Monte Agudo*,[48] but he was promptly attacked by the Almoravides and slain. Ibn Ḳasī was criticized by some of his devotees for his failure to supply aid on this occasion, but he excused himself with the explanation that this initial failure was but the false dawn that would inevitably be followed by the real.[49] And, indeed, his sun was about to rise. Though he felt unsafe and retired for a time to the seclusion of a farm in the district of Mértola, one of the bravest and astutest of his followers contrived with a band of seventy *Morīdīn* to seize the fortress of Mértola on the night of August 14th. The news of this remarkable success spread rapidly throughout al-Gharb and various other places joined the rebellion. By the beginning of September the movement had gone so far that Ibn Ḳasī was emboldened to enter Mértola at the head of his forces and install himself in the castle. He now assumed the pretentious title of *al-Mahdī* and made Mértola the capital of an independent state.[50]

Among the most important of Ibn Ḳasī's partisans were Abū Muḥammad Sīdrey Ibn Wazīr, who had led the revolt in Évora, and Abu'l-Walīd Muḥammad b. ʿOmar Ibn al-Mundhir of Silves, who had performed a like service in his home city and its environs.[51] They had coöperated together closely in the rising; and within a month after Ibn Ḳasī had established the seat of his government in Mértola they presented themselves before him and, recognizing him as their chief, they were confirmed in local governorships, the former over the district of Beja, the latter over Silves.[52] But for some reason relations of confidence were not maintained between Sīdrey Ibn Wazīr and Ibn Ḳasī, and so it was Ibn al-Mundhir who became the principal arm of the revolution.

[47] Codera, *op. cit.*, pp. 34–55.

[48] Of uncertain identification, though it appears to be mentioned in our text above, p. 634; cf. *ibid.*, note 338.

[49] Codera, *op. cit.*, pp. 36–37.

[50] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, pp. 199–200, 202; ʿAbd al-Wāḥid al-Marrākushī, *Histoire des Almohades*, French translation by E. Fagnan, p. 182; Codera, *op. cit.*, pp. 37–39.

[51] Among other exploits Ibn al-Mundhir stormed the castle of Monchique (apparently with the aid of forces supplied by Sīdrey Ibn Wazīr) and slew its defenders, thereby making such an impression on the Almoravides that they withdrew from Beja under a safe-conduct without a struggle: *ibid.*, p. 292; Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 202. For the identification of Monchique (Arabic *Morjīk*) see C. F. Seybold "*Monchique et Arrifana d'Algarve*," in *O Archeologo Português*, VIII (1903), 123–124.

[52] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, pp. 200, 202; Codera, *op. cit.*, pp. 39–40, 292–293.

Returning to his post in Silves and uniting his forces with those of the district of Ocsonoba, which had also joined the revolt, he crossed the Guadiana and promptly conquered both Huelva and Niebla. Then he advanced upon Seville, but being routed by Yaḥyā Ibn Ghaniyā, the leader of the Almoravide forces, he fell back precipitately on Niebla, and two days later he retired to Silves itself. However, Niebla continued to be held in the rebel interest and endured a siege during the winter of 1144–45, until pressure on it was relieved by a fresh rising against the Almoravides in Cordova.[53] Ibn Ḳasī now dispatched Ibn al-Mundhir with the forces of Silves and Niebla to seek recognition from the Cordovan rebels, but the expedition arrived too late. The Cordovans had already accepted other leadership. Then Ibn Ḳasī had to contend with an open revolt on the part of Sīdrey Ibn Wazīr, who had by this time extended his sway over Badajoz; and once more it was Ibn al-Mundhir who was called to action in the crisis. But the suppression of the revolt proved to be a task too great for him. His forces were defeated and he himself was taken and sent to Beja where, on orders from Sīdrey Ibn Wazīr, his eyes were dug out and he was held in close confinement until the coming of the Almohades procured his release.[54]

Deprived of the aid of his principal supporter, Ibn Ḳasī himself was doomed, though it is not clear how his collapse was brought about. He may have been induced to depart from his capital by means of some ruse,[55] perhaps after he had brought about a temporary recognition of the Almohades.[56] What seems more certain is that he was in some way overwhelmed and deprived of his capital by Sīdrey Ibn Wazīr in January or February, 1146 (Shaʿbān, 540 A.H.)[57] and that, having previously had some dealings with the Almohades, he crossed over to Morocco and invited the triumphant caliph ʿAbd al-Mu'min to send over an army and conquer al-Andalus.[58] The caliph accepted the invitation with alacrity. He sent Ibn Ḳasī back to Spain with an army under the command of Barrāz b. Muḥammad al-Masūfī, whom he directed to

[53] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, pp. 202–203; Codera, *op. cit.*, pp. 40–42.

[54] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 204; Codera, *op. cit.*, pp. 42–44.

[55] ʿAbd al-Wāhid al-Marrākushī, *Histoire des Almohades*, French translation by E. Fagnan, p. 182; cf. Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 200.

[56] Ibn al-Abbār, *loc. cit.;* Ibn Khaldūn *Histoire des Berbères*, ed. Casanova, II, 185.

[57] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 239; Ibn Khaldūn, *Histoire des Berbères*, ed. Casanova, II, 184.

[58] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 200; Ibn Khaldūn, *loc. cit.;* Codera, *op. cit.*, pp. 44–45.

make war on Almoravides and rebels alike.[59] The expedition met with immediate success. Tarifa, Algeciras, Jerez, and Niebla all submitted, and presently the invaders took the road to Mértola where, as has been noted, Ibn Ḳasī is said to have brought about an earlier recognition of the Almohades. Thence Barrāz proceeded against Silves, which was taken by assault and placed under the rule of Ibn Ḳasī. Then he moved upon Beja and Badajoz where Sīdrey Ibn Wazīr was brought to submission;[60] and so, having reduced the whole of the Southwest to subjection and united the rebel leaders under his command, Barrāz went on to the siege of Seville, which he took on January 17 or 18, 1147.[61]

The reward which Ibn Ḳasī received for the services which he had rendered the invaders was the governorship of his native province of Silves, but it proved inadequate to satisfy his restless ambition and insure his loyalty. When the tyrannical conduct of the Almohades provoked a fresh rising Ibn Ḳasī joined it.[62] He was promptly compelled to return to obedience by the new governor Yūsuf b. Sulaimān whom 'Abd al-Mu'min established in Seville, but, though he was not removed from his post in Silves, his loyalty was not won. When in A.H. 545 (A.D. 1150–51) the caliph summoned the chiefs of the Spanish Muslims to appear before him in Salé to swear fidelity and yield up their respective states, Ibn Ḳasī failed to answer the summons;[63] and resorting once more to open rebellion, he allied himself with Affonso Henriques, the Christian king of Portugal, who sent him rich presents. Such conduct enraged his subjects, prominent among whom was his former devoted supporter Ibn al-Mun<u>dh</u>ir, who, since the triumph of the Almohades, had been released from captivity and returned to the city over which he had formerly held sway. A local conspiracy was formed against Ibn Ḳasī, and in August or September 1151 (Jumādā I, 546 A.H.) he was overthrown and slain in the palace of al-<u>Sh</u>arājib and his head

[59] Ibn <u>Kh</u>aldūn, *loc. cit.;* Ibn al-Abbār, *loc. cit.;* 'Abd al-Wāḥid al-Marrāku<u>sh</u>ī, *Histoire des Almohades*, French translation by E. Fagnan, p. 182.

[60] By submitting Sīdrey was evidently able to remain in office and continue to exercise local authority. A contemporary Christian source records an attempted exchange of letters between him and the people of Lisbon during the Christian siege of that stronghold in the summer of 1147, perhaps in the month of August; and in both documents, in the inadequate mediaeval translation, he is entitled *Rex Eburensium* (king of Évora). See *De Expugnatione Lyxbonensi*, ed C. W. David, pp. 136–139.

[61] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, pp. 200, 239; Ibn <u>Kh</u>aldūn, *Histoire des Berbères*, ed. Casanova, II, 185; Codera, *op. cit.*, pp. 45–47. According to Ibn al-Abbār both Sīdrey Ibn Wazīr and Ibn Kasī were present at the siege of Seville. In general on the Almohade conquest of al-Andalus and the sources therefor see E. Lévi-Provençal, *Documents Inédits d'Histoire Almohade* (Paris, 1928, being *Textes Arabes Relatifs à l'Histoire de l'Occident Musulman*, I), p. 202, note 3.

[62] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 200; Ibn <u>Kh</u>aldūn, *Histoire des Berbères*, ed. Casanova, II, 186.

[63] Ibn <u>Kh</u>aldūn, *op. cit.*, II, 188.

was impaled on a lance which he had received as a present from the Portuguese monarch; and the blind Ibn al-Mundhir ruled in his stead.[64]

What happened thereafter is not entirely clear. Apparently Ibn al-Mundhir, who had become a true convert to the teachings of the Mahdī Ibn Tūmart, ruled Silves for a time in due subjection to the Almohades; but presently (the exact date is uncertain) he was thrust aside by the ambitious Sīdrey Ibn Wazīr.[65] However, Sīdrey was not left in control for long. With a view to strengthening the position of the Almohades in al-Andalus, ʻAbd al-Mu'min established his son Abū Yaʻḳūb Yūsuf as governor in Seville, and the latter, setting out vigorously to suppress independence or rebellion on every hand, soon possessed himself of the whole region over which Sīdrey Ibn Wazīr held sway; and so Silves was brought under the direct rule of the Almohades in the year of the Hijra 552 (13 February 1157—3 January 1158 A.D.).[66]

Thereafter it may be presumed that Silves remained in due subjection to the Almohade empire until disaster overtook it at the hands of crusaders and the Portuguese in 1189.[67] The intermittent warfare of the Almohades and the Portuguese was particularly violent at times, the raids of each carrying far into the territory of the other,[68] but neither side gained a decisive or permanent advantage; and Silves seems to have lain too far to one side to become directly involved, though it contributed artisans for the rebuilding of Beja in 1175[69] after its destruction by Affonso Henriques, and the naval warfare between the fleets of Ceuta and Seville on the one side and Lisbon on the other swept before it without important result from 1179 to 1181.[70] The death of the caliph Abū Yaʻḳūb Yūsuf in 1184 during a famous raid into Portugal which threatened Santarém and carried as far as Tôrres Vedras and Alcobaça[71] did not seriously alter the situation. His son Abū Yūsuf Yaʻḳūb al-Manṣūr succeeded him without opposition and ruled with equal or greater might. Though he was preoccupied with revolt in Africa for

[64] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, pp. 200, 204.

[65] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, pp. 204, 239. Ibn al-Mundhir was transferred to Seville for a time and then crossed over to Morocco. He died at Salé in A.H. 558 (A.D. 1162–63).

[66] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, pp. 204, 239; Ibn Khaldūn, *Histoire des Berbères*, ed. Casanova, II, 192; Codera, *op. cit.*, pp. 51–52.

[67] On the general position of the Almohades and the Muslim opposition which they encountered see Alfred Bel, *Les Benou Ghânya, Derniers Représentants de l'Empire Almoravide, et Leur Lutte contre l'Empire Almohade* (Paris, 1903), pp. 20–22.

[68] *El Anónimo de Madrid y Copenhague*, ed. Huici, pp. 3–11, 17, 19, 32–37; Ibn Khaldūn, *op. cit.*, II, 198, 204.

[69] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 10.

[70] *Ibid.*, pp. 15–16, 19; cf. Ibn Khaldūn, *op. cit.*, II, 202, 204.

[71] See the full analysis with the sources in R. Dozy, *Recherches sur l'Histoire et la Littérature de l'Espagne pendant le Moyen Âge*, II, 443–480.

a number of years,[72] his position in Spain was not thereby seriously weakened, and by 1188 he was planning a renewal of the holy war.[73] He was competently represented in Seville by one of his kinsmen,[74] and Silves was being governed by Aīsa b. Abū Ḥafṣ b. ʿAlī who was apparently a son of the great shaikh Abū Ḥafṣ ʿOmar and is described as "greatly experienced in the defense of frontiers." [75] There was no Muslim rebellion in any quarter, and the menace of the Christians did not appear to be particularly acute.

And yet, as it happened, the Christian menace suddenly became grave in the summer of 1189 before al-Manṣūr was prepared to meet it. Alfonso VIII of Castile had carried a raid throughout a good part of Muslim Spain. His columns had swept through the region of Cordova and over the fair and fertile Aljarafe to the east of Seville. A number of important castles had been seized, and, according to the *Anales Toledanos*, Alfonso's forces had even reached the sea.[76] "Al-Andalus," says a Muslim contemporary, "swayed back and forth with encounters and surprises, and the inhabitants were like a weaver's shuttle." [77] And while the governor of Seville was thus fully occupied with the Castilian menace and unable to turn his attention elsewhere, the blow had fallen on the shores of distant al-Gharb. Alvor had been stormed and sacked with an appalling slaughter in June,[78] and in July, August, and September Silves had suffered a fate hardly less terrible; and there had been none but local forces to offer resistance.[79]

But the Christian triumph was to be shortlived for the news of the disasters of 1189 stirred al-Manṣūr to an extraordinary effort to redress the balance. Sending word to his representatives in Spain to make extraordinary preparations,[80] he himself crossed the Strait with a large army in the spring of 1190 [81] and advanced from Tarifa on June 6.[82]

[72] *El Anónimo de Madrid y Copenhague*, ed., Huici, pp. 41–58; Ibn Khaldūn, *op. cit.*, II, 206–212; Bel, *Les Benou Ghânya*, pp. 28–88.

[73] *El Anónimo de Madrid y Copenhague*, ed., Huici, p. 60.

[74] Abū Ḥafṣ Yaʿkūb, son of his uncle Abū Ḥafṣ, according to *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 60, 62; Abū Yūsuf Ibn Hafṣ, according to Ibn Khaldūn, *op. cit.*, II, 212.

[75] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 61; see above, p. 628, note 282. He is mentoned later by Ibn Khaldūn (*op. cit.*, II, 215) as a witness of al-Manṣūr's political testament.

[76] For further details see *El Anónimo de Madrid y Copenhague*, ed. Huici, pp. 61–62; *Anales Toledanos*, in Flórez, *España Sagrada*, XXIII, 393; cf. A. Ballesteros y Beretta, *Historia de España* (Barcelona, 1918–36), II, 266–267; Ibn Khaldūn, *op. cit.*, II, 212.

[77] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 62.

[78] See below, Appendix B, pp. 663–666.

[79] According to Ibn Abī Zarʿ (*al-Ḳirṭās*, being *Annales Regum Mauritaniae*, ed. C. J. Tornberg, II, 191, 192), whose information is certainly to some extent confused and of doubtful value, Beja was also taken by the Chirstians, and lost again in 1191 with the loss of Silves.

[80] Cf. *El Anónimo de Madrid y Copenhague*, ed. Huici, pp. 60, 63–64.

[81] *Ibid.*, p. 63: "domingo 23 de Rabía." Since the author does not say whether he means Rabīʿ I or Rabīʿ II, and since the 23rd was not a Sunday in either month in this year, it is

He ordered the governor of Seville to lead a large force to the siege of Silves, and by early July the place had been invested and was under attack both by land and sea [83]—but not, as it would seem, before the Portuguese garrison had been considerably strengthened by the chance arrival of a storm-tossed vessel from London bearing more than eighty sturdy crusaders who agreed to remain and participate in the defense of the city.[84] Meanwhile, the caliph himself struck at the heart of King Sancho's dominions. Advancing with a large force which had been assembled and equipped in Cordova and Seville, he crossed the Tagus on June 24, took Tôrres Novas after a short siege, and pushed on to beleaguer the great Templar stronghold of Tomar.[85] The danger to the Portuguese monarchy seemed extreme, and our Almohade authority asserts that al-Manṣūr refused the peace for which Sancho sued and ordered the devastation of the country as far as Coimbra.[86] But the danger evidently was not so great as it appeared to be, and the timely arrival in Santarém of a band of English crusaders who had recently landed at Lisbon suddenly turned the scale.[87] Our Christian authority declares that it was al-Manṣūr who asked for peace and Sancho who refused it.[88] In any case the caliph fell back precipitately on Seville, where he arrived on July 16, and ordered the lifting of the siege of Silves.[89] The reason assigned in the Muslim source for this sudden suspension of hostilities—"porque los precios habían subido y los socoros del país comenzaban a faltar"—is not unnatural in this kind of

impossible to give an exact date. Huici has erred in his inference that the date should be April 30.

[82] *El Anónimo de Madrid y Copenhague,* ed. Huici, p. 64.

[83] *Ibid.*, 64: "Salió el Sid con todos los soldados de que hemos hecho mención, después de reunir a los que se habían adelantado, con los que se habían retrasado, el primero de Chumada el aual [June 6] y continuaron su marcha hasta llegar delante de Silves. A fines del citado mes [Jumādā I ended July 5] llegaron las escuadras, multiplicáronse las algaradas por tierra y por mar; levantáronse máquinas y aprestos de guerra; acercáronse al muro las tropas de línea y apretaron a los sitiados, hasta ahogarlos y, entretanto que continuaba el cerco, arreciaba unas veces el ataque y se calmaba otras."

[84] For full details see *Gesta Regis Henrici Secundi,* ed. Stubbs, II, 116–118.

[85] *El Anónimo de Madrid y Copenhague,* ed. Huici, pp. 64–65 ("el río de Beja" in the text is probably an error for "el río Tajo"); *Gesta Regis Henrici Secundi,* ed. Stubbs, II, 118; Roger of Howden, *Chronica,* ed. Stubbs, III, 175; Ibn Khaldūn, *op. cit.*, II, 212.

[86] *El Anónimo de Madrid y Copenhague,* ed. Huici, pp. 65–66.

[87] *Gesta Regis Henrici Secundi,* ed. Stubbs, II, 118–119.

[88] *Ibid.:* "Qui [al-Mansūr] cum audisset adventum peregrinorum timuit valde, et missis legatis suis ad regem Portingalensen, petiit ab eo Silviam civitatem, et ipse recederet, et redderet ei castellum [Tôrres Novas] quod ceperat, et pacem haberet cum illo septem annis."

[89] *El Anónimo de Madrid y Copenhague,* ed. Huici, p. 66; *Gesta Regis Henrici Secundi,* ed. Stubbs, II, 119. According to the latter source the precipitate retirement was explained to the Christians by a false report of the caliph's death. I have not discovered on what authority Herculano (*Historia de Portugal,* 8th ed., III, 206) dates the lifting of the siege of Tomar on July 11. Tôrres Novas was evidently abandoned.

warfare in which the army was expected to live off the country,[90] yet it can hardly be doubted that the arrival of the northern crusaders was also an important consideration.[91]

But al-Manṣūr's abandonment of the struggle was only temporary. Next year he prepared himself once more and renewed the contest. Setting out from Seville at the end of April, he took Alcácer do Sal on or about June 10 after an elaborate siege in which he was assisted by a fleet in the Sado.[92] He ordered the fortress to be repaired, and placed it under the command of the experienced general and poet, Abū Bakr Muḥammad b. Sīdrey Ibn Wazīr al-Shilbī (of Silves),[93] son of that Abū Muḥammad Sīdrey Ibn Wazīr who had played so prominent a rôle, first as a supporter and then as an enemy of Ibn Ḳasī, in the previous generation.[94] Then pushing on, he forced the capitulation of the lofty stronghold of Palmela, and ordered its destruction.[95] Then, advancing still further, he reached the shores of the Tagus opposite Lisbon and took the castle of Almada.[96] Thereafter he turned southward and, arriving before Silves on June 27, he invested it so closely that all communication with the outside world was effectively cut and the garrison felt its situation to be hopeless. On July 10 a truce was arranged to give time for communication with King Sancho, and permission to capitulate having been obtained, the Christians marched out some ten days later.[97] Leaving Silves on July 23, al-Manṣūr returned to Seville

[90] Cf. E. Lévi-Provençal, *L'Espagne Musulmane au Xème Siècle*, p. 139: "Il fallait, pour que le calife décidât la mise sur pied d'une expédition estivale, que la récolte s'annonçât belle: comme l'armée se ravitaillait sur place, cette condition était essentielle. . . . Les années de disette, bien entendu, on ne pouvait songer à partir en guerre."

[91] Indeed, in the light of what had happened in 1189 and of the knowledge which al-Manṣūr must have had of the general outpouring of maritime crusaders from the North, he may well have considered that a prolonged siege of Silves was not to be thought of until after the northerners had passed.

[92] *El Anónimo de Madrid y Copenhague*, ed. Huici, pp. 68–69; Roger of Howden, *Chronica*, ed. Stubbs, III, 175.

[93] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 69; Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 239; cf. al-Makkarī, in Pascual de Gayangos, *History of the Mohammedan Dynasties in Spain*, II, 320; Ibn Khaldūn, *Histoire des Berbères*, ed. Casanova, II, 212. His full name as given by Ibn al-Abbār was Abū Bakr Muḥammad b. Sīdrey Ibn 'Abd al-Wahab Ibn Wazīr al-Keisi; al-Maḳḳarī calls him Abū 'Abd Allāh Ibn Wazīr al-Shilbī. See below, p. 662.

[94] Ibn al-Abbār, *loc. cit.;* compare above, pp. 653–656.

[95] *El Anónimo de Madrid y Copenhague*, ed. Huici, pp. 69–70; Roger of Howden, *Chronica*, ed. Stubbs, III, 175. The garrison of Palmela was permitted to retire into the unconquered territory of King Sancho.

[96] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 70. Lacunae in the text make it impossible to determine from this source whether Almada was taken or not, but Roger of Howden (*op. cit.*, III, 175) puts the matter beyond doubt.

[97] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 70: "Entonces levantó el campo y se dirigió, con la ayuda de Dios, hacia Silves, a donde llegó el jueves 2 de Chumada el ajer. . . . Rodeáronla los campamentos por todas partes, de modo que no le dejaban respirar, ni podían llegar a ella noticias de cristianos por ningún camino. Se colmaron los fosos de escombros,

on the 28th and so concluded a three months' campaign of continuous victories; [98] and early in October he crossed over to Morocco.[99]

The Portuguese attempt to dominate a large part of the region beyond the Tagus had collapsed and for a good many years to come it seems to have been the deliberate policy of the monarchy not to renew it.[100] Meanwhile the power of the Almohades in Spain remained strong and with the crowning victory of Alarcos in the summer of 1195 it became almost overwhelming. And yet so distant and exposed a post as Silves remained open to sudden attack; and, if we can accept the unsupported authority of the English chronicler, Roger of Howden, it was seized and greatly damaged by a band of German maritime crusaders in 1197, though his assertion that they completely destroyed it is doubtless an exaggeration.[101] In any case the power of the Almohades in al-Gharb cannot have suffered more than a temporary shock, and the surviving evidence shows that their orderly administration remained intact for some years longer. The *sayyid* (lord) Abū 'Abd Allāh b. Abū Ḥafṣ b. 'Abd al-Mu'min was appointed to rule the province in 1198,[102] and the *sayyid* Abū Muḥammad 'Abd al-Wāḥid b. Abū Ya'ḳūb

y la desgracia los visitaba con sus rayos mañana y tarde. El miércoles 15 del mes . . . se aprovecharon de la mañana los cristianos para reunirse; confiaron en que aquella hora no era propia para el ataque; los musulmanes expiaban los azares de la guerra con la vigilancia de la luna nueva de Xaual. Conoció uno de los adalides la negligencia de los cristianos y el descuido y el sueño a que estaban entregados; asaltó el muro, y acudió tras él un grupo de soldados que levantaron banderas, redoblaron los tambores y se llenó el aire de gritos invocando a Dios y de voces. Los cristianos no se despertaron hasta verse en manos de la muerte, entre heridos y batidos; se encontraron en un lago de su propia sangre, y clamaron pidiendo con ansia una capitulación; concedióles un plazo de diez días y se separaron. Su tirano permitióles capitular, y les dió las gracias por su constancia en tan grande prueba; salieron del castillo de Silves, el jueves 23 de Chumada el ajer." The last date is inconsistent, Thursday being the 23rd in this year. See also 'Abd al-Wāḥid al-Marrākushī, *Histoire des Almohades*, French translation by E. Fagnan, p. 243; Ibn Khaldūn, *op. cit.*, II, 212; Ibn al-Athīr, *Annales du Maghreb et de l'Espagne*, French translation by E. Fagnan, pp. 608–609.

As indicated above (p. 657, note 79), according to Ibn Abī Zar', Beja was lost by the Christians at this time, but I am inclined to believe they were not holding it.

[98] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 70.

[99] *Ibid.*, p. 71.

[100] It is to be observed that Silves capitulated to the Almohades in 1191 with the consent of King Sancho and, as far as we know, he displayed no interest in the German attack on the place in 1197. His successor, Affonso II, appears to have taken no part in the conquest of Alcácer do Sal by northern crusaders in 1217. See Herculano, *Historia de Portugal*, 8th ed., IV, 95–98.

[101] *Chronica*, ed. Stubbs, IV, 26: "Exercitus autem imperatoris, qui de Alemannia et caeteris terris ejus iter Jerosolimitanum per mare susceperat, transitum fecit in Normanniam et Angliam; et sic rectum cursum tenens usque in Hispaniam, eripuit civitatem Silviae de manu paganorum; quam penitus destruxerunt, non relinquentes lapidem super lapidem: timebant enim quod, si eam tradidissent regi Portugalensi, ipse eam amitteret, sicut prius fecerat." Though Roger of Howden is the only chronicler who mentions this attack upon Silves, the German maritime crusade is noted briefly in *Chronica Regia Coloniensis* (ed. Georg Waitz, Hanover, 1880, p. 160) and in *Annales Stadenses* (*M. G. H.*, *Scriptores*, XVI, 353).

[102] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 89; Ibn Khaldūn, *op. cit.*, II, 215.

was made governor of "Silves and al-Gharb" in 1203.[103] As late as 1210 or 1211 (A.H. 607) we note the appointment of a kāḍī of Silves, the kā'id Abū 'Abd Addār Aīsa of Murcia.[104]

But during the period of Almohade decline after the disaster of Las Navas da Tolosa (1212) we lose track of events in the far Southwest, and the final conquest of al-Gharb by the Portuguese in 1249 finds no echo in the Muslim sources and remains an event of much obscurity in the Christian.[105]

Idrīsī's praise of Silves as a centre of culture [106] is reëchoed in the works of later geographers, notably Abū'l-Fidā' [107] and al-Kazwīnī,[108] and it finds considerable support in concrete facts which may be briefly enumerated. As noted above, the well-known poet of the eleventh century Ibn 'Ammār, whose life, both political and literary, is so closely connected with Silves, was a native of the province; and he found competent instruction and completed a good part of his education under the grammarian of Silves, Abu'l-Hajjāj Yūsuf b. 'Isā al-A'lam.[109] The poet king of Seville, al-Mu'tamid, whose verses on Silves have already been mentioned,[110] and who was for a time the city's governor, is declared by Abu'l-Fidā' to have been brought up there.[111] 'Abd al-Wāḥid al-Marrākushī, who relates how Ibn 'Ammār at a critical moment in his career found patronage at the hands of a wealthy merchant of Silves,[112] tells a similar story of the twelfth-century poet Ibn H'abbūs who had the good fortune to fall into the hands of an inhabitant of Silves who had set aside the income of a piece of land worth a hundred dinars a year for the support of poets.[113] The devotion to poetry of the twelfth-century rebels whose names are associated with Silves seems remarkable. Poems have survived by Ibn Ḳasī himself,[114] as well as by his follower Ibn al-Mundhir,[115] and by a certain Abū 'Omar Aḥmad b. 'Abd Allah b. Ḥaryūn al-Shilbī (of Silves), who is described as one of Ibn Ḳasī's ministers; [116] and while we hear nothing of verse-making by

[103] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 98.
[104] *Ibid.*, p. 114.
[105] See Herculano, *Historia de Portugal*, 8th ed., V, 88–92.
[106] Above, pp. 645–646.
[107] *Géographie d'Aboulféda*, French translation by [J. T.] Reinaud, II, 237.
[108] Cited by R. A. Nicholson, *A Literary History of the Arabs*, London, 1907, p. 416.
[109] See above, p. 650, note 33.
[110] See above, p. 651.
[111] *Géographie d'Aboulféda*, French translation by [J. T.] Reinaud, II, 237, 240.
[112] *Histoire des Almohades*, French translation by E. Fagnan, pp. 97–98.
[113] *Ibid.*, pp. 184–185.
[114] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 201. Ibn Ḳasī is also credited with the authorship of a book entitled *Khal 'al-Na'lain fī 'l-Tasạwwuf.* See *Encyclopaedia of Islam*, *s. v. Ibn Ḳasī.*
[115] Ibn al-Abbār, in Dozy, *Notices sur Quelques Manuscrits Arabes*, p. 204.
[116] *Ibid.*, p. 201.

Abū Muḥammad Sīdrey Ibn Wazīr, we do have poems by his son Abū Bakr Ibn Wazīr al-S̲h̲ilbī whose reputation as it survived in the later work of al-Makkarī was quite as illustrious for poetry as for generalship.[117] In another passage al-Maḳḳarī tells us of a famous man of letters, Abū Merwān ʿAbd al-Malik Ibn Bedrūn, and of a famous grammarian, Abū Muḥammad ʿAbd Allāh, son of Ibn al-Sīd of Badajoz, both of whom were born in Silves.[118] Finally, Dozy refers to Ibn Ḳāsim, a historian of the reign of al-Muʿtamid, who was a native of Silves.[119]

[117] *Ibid.*, p. 239; al-Maḳḳarī, in Pascual de Gayangos, *History of the Mohammedan Dynasties in Spain*, II, 320; compare above, p. 659.

[118] Pascual de Gayangos, *History of the Mohammedan Dynasties in Spain*, I, 62.

[119] Dozy, *Histoire des Musulmans d'Espagne*, new ed., III, 181.

APPENDIX B

THE CONQUEST OF ALVOR (1189) AND THE CONFUSION OF THE SOURCES DEALING THEREWITH

For the conquest and destruction of Alvor the Turin text published in the present work is by far the best surviving source:[1] the author wrote from good information and he indicates where he got it. Ships from Germany and from Flanders (*de nostro imperio et de Flandria*) had preceded by four or five weeks the fleet which, bearing the author, arrived in the port of Lisbon on July 3 or 4: that is to say, this earlier fleet must have been in the Tagus at the end of May or the beginning of June. Thereafter it had sailed on beyond Lisbon and stormed the stronghold of Alvor (*castrum nomine Alvor*), a dependency of Silves; and there had been time for a Portuguese galley to accompany it beyond Alvor as far as the Strait of Gibraltar and to return to Lisbon bearing Saracen prisoners and the good news before the departure of the author of the Turin text on July 14 for the further voyage to Silves. Evidently the conquest of Alvor had occurred during the month of June, as Kurth (pp. 164, 174) has calculated; a more definite date cannot be given.

The storming of Alvor culminated in a frightful slaughter of the inhabitants of the port and other neighboring communities which made a deep impression on our author. Neither age nor sex had been spared, he says, and according to reports which he believed veracious, about 5,600 had been done to death; and, reporting the passing of the fleet on which he himself sailed, he adds, "We saw Alvor Castle in ruins, situated on the sea, and other deserted castles whose inhabitants had been killed in Alvor"; and in a later passage, where he gives a list of the *castella* which fell into the hands of the Christians as a result of the conquest of Silves, he remarks on their emptiness, a part of their inhabitants having been killed in Alvor though the greater part had fled into Silves.

The frightful massacre of the Muslims at Alvor evidently finds an echo in the one Arabic source which seems to refer to the event, namely the Anonymous of Madrid and Copenhagen. The author of this remarkable contemporary work concludes his narrative of the Christian conquest of Silves in 1189 as follows: "Destruyó el enemigo un castillo de la comarca, llamado del Puerto, y dió muerte a cuantos en él había, grandes y pequeños, hombres y mujeres. Dios premie su martirio el diá de la resurrección."[2] This account, coming just after the fall of Silves,

[1] Above, pp. 616–617, 634.

[2] *El Anónimo de Madrid y Copenhague*, ed. Huici, p. 61.

is somewhat out of order in the narrative; and yet it seems all but certain that by *El Puerto* the author refers to Alvor, the *Portus Hannibalis* of the ancients.

Apart from the Turin text, the only contemporary, or nearly contemporary, Christian work to refer specifically to the conquest of Alvor seems to be the *Chronica Regia Coloniensis*, and here again there is evidence of the deep impression made by the ruthless slaughter of the inhabitants. The account, with some abridgment, is as follows: "In quadragesima [1189] naves undelibet adventantes et sibi invicem copulatae velis oppansis iter aequoreum ingressae sunt. Quae post decem dies navigationis suae prosperante Deo terrae Sancti Jacobi applicuerunt. Erant sexaginta naves ex eis, virorum vero pugnatorum ad 10 milia et amplius . . . Inde . . . cum in Affricae partes venissent, urbem gentilium populosam nomine Albur oppugnant et capiunt atque aurum et argentum infinitum inde detrahunt, urbicis in ore gladii cesis." [3]

Though it is impossible to be quite certain, it seems very probable that it is this same expedition (rather than the later one which resulted in the conquest of Silves) which is described by the contemporary annalist Lambertus Parvus (d. 1194), a monk of Liége, who tells of a force of Frisians, Danes, Flemings, and men of Cologne and Liége who set sail in fifty-five ships in the spring of 1189 and waged many battles with the pagans as they passed around Spain before going on to Sicily and to Acre.[4]

Lambert makes no mention either of Alvor or the massacre; but an echo of the latter, now erroneously transferred from Alvor to Silves, surely reaches us in another contemporary or nearly contemporary, but confused, narrative which has survived in almost identical words in several places. It may be quoted from the *Chronicon* of Robert of Auxerre, where it is likely to be in as nearly its original form as it is now possible to obtain it: "De Fresia et Datia 50 naves pariter federate

[3] *Chronica Regia Coloniensis*, ed. Waitz, pp. 142–143.

[4] Lambertus Parvus, *Annales*, in *M.G.H.*, *Scriptores*, XVI, 649. The arrival at Acre on September 1 of the naval forces which had destroyed Alvor is perhaps reported by Arnold of Lübeck (*Chronica Slavorum*, ed. J. M. Lappenberg, Hanover, 1868, pp. 141–142) as follows: "Denique tertia die obsidionis in Kalendis Septembris apparuit multitudo navium de diversis Teutonicorum partibus venientium. . . . Fuerunt autem quinquaginta quinque naves Teutonicorum expansis velis applicantes." Their arrival is also mentioned by Ralph of Coggeshall (*De Expugnatione Terrae Sanctae Libellus*, in his *Chronicon Anglicanum*, ed. Joseph Stevenson, London, 1875, p. 252); and the *Itinerarium Peregrinorum et Gesta Regis Ricardi* (ed. William Stubbs, London, 1864), p. 65, after recording the same event, erroneously accredits them with the conquest of Silves (instead Alvor): "Nec praetereundum silentio quod a praetereuntibus insigniter est gestum, nam urbem quandam in maritima Hispaniae quae Silvia dicitur, audaciter impetunt, citius expugnant, caesisque Gentilibus incolis et urbe Christianis tradita, ibidem etiam ordinato antistite, victores procedunt."

eandem peregrinationem arripiunt. De Flandriis quoque 37 rates cum grandi apparatu secute sunt, et dum per Hispanias transeunt, quandam Saracenorum urbem nomine Silviam obsident. Et post 40 dies captam diripiunt. Nulli etati parcitur, sexus uterque pariter trucidatur. Plurimum in ea opum et ornamentorum repertum est et inter eos equa distributione partitum. Regi de Portugal urbem reliquere tenendam. Alia quoque Saracenorum oppida ab eis direpta sunt." [5] From Robert of Auxerre this account was embodied in the chronicle of William of Nangis [6] and also in that of St. Martin's of Tours; [7] and according to a Portuguese scholar of the eighteenth century, Vincente Salgado, it also appeared in a manuscript chronicle (which he ascribed to a certain "Hugo, author coevo daquelles dias") in the library of the bishop of Beja.[8] Verbal similarities can leave little doubt that this narrative is also the ultimate source of a letter purporting to be from Pope Clement III to the Byzantine emperor Isaac Angelus (reporting *inter alia* the conquest of Silves) which has often been cited as genuine but which was in fact a crude fabrication of the sixteenth century.[9]

As Kurth (p. 169) rightly perceived, there is in this narrative a confusion of two major enterprises. The fifty ships from Frisia and Denmark must pretty certainly be assigned to the earlier expedition which resulted in the conquest of Alvor, as must also the reference to the slaughter of the defeated: there was no such massacre at the conquest of Silves. On the other hand we know that the northern force which sailed to the conquest of Silves consisted of thirty-seven vessels, though of course they were not all Flemish; and it is evident that the statement about a siege of forty days, while not strictly accurate, must refer to

[5] Robert of Auxerre, *Chronicon*, in *M.G.H.*, *Scriptores*, XXVI, 254.

[6] *Chronicon*, in *Recueil des Historiens . . . de la France* (ed. Martin Bouquet and others, Paris, 1738–1904), XX, 745.

[7] *Chronicon Turonense*, in E. Martène and U. Durand, *Veterum Scriptorum . . . Amplissima Collectio* (Paris, 1724–33), V, col. 1031–1032.

[8] Vincente Salgado, *Memorias Ecclesiasticas do Reino do Algarve* (Lisbon, 1786), I, 268; cf. Kurth, p. 168, note 9.

[9] It is undated but was assigned by Philipp Jaffé (*Regesta Pontificum Romanorum*, 2nd ed., Leipzig, 1888, no. 16373), quite impossibly, to the end of 1188. It was cited as genuine by Riant in his *Expéditions et Pèlerinages des Scandinaves en Terre Sainte au Temps des Croisades*, (1865), p. 277, though at a later date he was undoubtedly aware of its spurious character. It has also been cited as genuine, among others, by R. Röhricht (*Beiträge zur Geschichte der Kreuzzüge*, II, 200, note 122), Kurth (p. 168), and Chroust (p. cii). Attention was drawn to its true character by Johannes Geyer (*Papst Klemens III, 1187–1191*, Bonn, 1914, being *Jenaer Historische Arbeiten*, VII, 36). Published in Nicolaus Reusner, *Epistolarum Turcicarum . . . Libri V* (Frankfort, 1598–1600), pp. 16–17, it evidently was taken from *Epistolae Principum, Rerumpublicarum, ac Sapientum Virorum, ex Antiquis & Recentioribus, tàm Graecis, quàm Latinis Historiis & Annalibus Collectae* (Venice: *apud Jordanum Zilettum*, 1574), p. 131, "un recueil d'exercises épistolaires" made by Jeronimo Doncellini, an obscure physician of Verona. See Paul Riant, "Inventaire Critique des Lettres Historiques des Croisades," in *Archives de l'Orient Latin* (Paris, 1881–84), I, 126.

Silves, since Alvor was probably taken by storm in a very short time and, in any case, its conquest cannot have required forty days.[10]

The number and variety of references to northern maritime crusaders sailing for the Mediterranean during 1189, and perhaps also before the end of 1188, are very confusing. Kurth's effort (pp. 161–183) to fit them all together and bring them into harmony, while interesting and helpful, is not wholly convincing, particularly in view of the fact that he makes no allowance for sailings of which no record has survived. A careful reading of the extant sources leaves the impression that while two distinct international expeditions are to be reckoned with in the summer of 1189 (namely, the earlier which stormed Alvor and the later which conquered Silves), there was a more or less continuous outpouring of northern *maritimi* during the sailing season and that some sailings have gone unrecorded. Attention may be drawn to a statement by the author of the Turin text, made in connection with his enforced delay of twenty-three days at Sandwich: "Ibi et alie [naves] ad nos venerunt, sed pro diversis necessitatibus quedam precesserunt, quedam tardius subsequute sunt." [11] A statement of the Anonymous of Madrid and Copenhagen, made apropos of the Christian conquest of Silves in 1189, is also worth quoting in this connection: "Ocurrió la llegada de cierto número de naves cristianas, que iban según su costumbre a Jerusalén, desde que había sido arrancada a la gente de su religión; iban cada año para cumplir un voto que hacían, según sus creencias religiosas, y librarse de la promesa y de las condiciones a que los habían sujetado sus religiosos." [12]

Still another passage in the same source [13] gives an account of a Moorish naval victory, apparently in the region of the Strait of Gibraltar, which resulted in the destruction of Christian ships and the death of a part of their crews and the captivity of the rest—a victory of such importance as to cause the name of al-Manṣūr to be celebrated by the poets. The account leaves an unfortunate uncertainty concerning the exact date of this naval engagement, but it would seem to have occurred in the early spring of 1190, a date at which one would hardly have expected northern vessels to be in the Strait, unless they had wintered in some Portuguese harbor such as Lisbon. In any case we have evidence here of Christian ships endeavoring to pass the Strait of which there seems to be no record in Christian sources.

[10] Above, pp. 616–617.
[11] Above, p. 611.
[12] *El Anónimo de Madrid y Copenhague*, ed. A. Huici, pp. 60–61.
[13] *Ibid.*, p. 63.

The Iberian Peninsula and Northwest Africa to Illustrate the *Narratio de Itinere Navali.*

INDEX.

Mediaeval names of persons are arranged alphabetically under the English form of the Christian name, mediaeval place names alphabetically under the modern form. Bibliographical citations are, with rare exceptions, to the first or main reference only.

INDEX TO VOLUME 81